



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA
DOUTORADO EM ENGENHARIA ELÉTRICA

ERICK COSTA BEZERRA

**ABORDAGEM AUTO-ADAPTATIVA BASEADA NO CONCEITO DE EXPECTATIVA
DE VIDA APLICADA AOS MÉTODOS *PARTICLE SWARM OPTIMIZATION* E
MÁQUINAS *KERNEL* PARA PREVISÃO DA VELOCIDADE DO VENTO E GERAÇÃO
EÓLICA**

FORTALEZA

2022

ERICK COSTA BEZERRA

ABORDAGEM AUTO-ADAPTATIVA BASEADA NO CONCEITO DE EXPECTATIVA DE
VIDA APLICADA AOS MÉTODOS *PARTICLE SWARM OPTIMIZATION* E MÁQUINAS
KERNEL PARA PREVISÃO DA VELOCIDADE DO VENTO E GERAÇÃO EÓLICA

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia Elétrica. Área de Concentração: Energias Renováveis e Sistemas Elétricos

Orientadora: Prof^a. Ph.D. Ruth Pastôra
Saraiva Leão

Coorientador: Prof. Dr. Arthur Plínio de
Souza Braga

FORTALEZA

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- B469a Bezerra, Erick Costa.
Abordagem auto-adaptativa baseada no conceito de expectativa de vida aplicada aos métodos Particle Swarm Optimization e máquinas kernel para previsão da velocidade do vento e geração eólica / Erick Costa Bezerra. – 2022.
135 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, Fortaleza, 2022.
Orientação: Profa. Dra. Ruth Pastôra Saraiva Leão.
Coorientação: Prof. Dr. Arthur Plínio de Souza Braga.
1. Previsão. 2. Velocidade do vento. 3. Geração eólica. 4. Otimização por enxames de partículas. 5. Máquinas kernel. I. Título.

ERICK COSTA BEZERRA

ABORDAGEM AUTO-ADAPTATIVA BASEADA NO CONCEITO DE EXPECTATIVA DE VIDA APLICADA AOS MÉTODOS *PARTICLE SWARM OPTIMIZATION* E MÁQUINAS *KERNEL* PARA PREVISÃO DA VELOCIDADE DO VENTO E GERAÇÃO EÓLICA

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia Elétrica. Área de Concentração: Energias Renováveis e Sistemas Elétricos

Aprovada em: 05 de Maio de 2022

BANCA EXAMINADORA

Prof^a. Ph.D. Ruth Pastôra Saraiva Leão (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Arthur Plínio de Souza Braga (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Guilherme de Alencar Barreto
Universidade Federal do Ceará (UFC)

Prof. Dr. Hugo Valadares Siqueira
Universidade Tecnológica Federal do Paraná (UTFPR)

Prof. Dr. Alexandre Carlos Araújo da Costa
Universidade Federal de Pernambuco (UFPE)

À minha família. Tudo que consegui só foi possível graças ao amor, apoio e dedicação que vocês sempre tiveram por mim.

AGRADECIMENTOS

À minha orientadora, Prof^a. Ruth Pastôra Saraiva Leão, Ph.D., ao meu co-orientador, Prof. Arthur Plínio de Souza Braga, Dr., e ao meu orientador no exterior, Prof. Pierre Pinson, Dr., pelo incentivo na realização deste trabalho e pela paciência, compreensão e direcionamento. A vocês minha admiração e agradecimentos.

À CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro necessário à realização desse trabalho e desenvolvimento científico.

Ao Eng^o. Sérgio Armando Benevides Filho, Me, que sem os históricos de medição seria impossível desenvolver a primeira etapa do trabalho.

Ao Prof. Lars Kai Hansen, Dr., que disponibilizou o acesso aos *clusters* computacionais para a execução da pesquisa.

Aos professores e todos do Departamento de Engenharia Elétrica da UFC, responsáveis diretamente ou indiretamente pela minha formação no programa de doutorado.

*“Forecasting is very difficult, especially if it’s
about the future...”*

(Niels Bohr)

RESUMO

O aumento do uso da geração eólica tem apresentado desafios ao planejamento e operação dos sistemas de potência e mercado de energia, sendo necessária uma ferramenta de auxílio à tomada de decisão capaz de lidar com sua intermitência. Para este propósito, uma abordagem auto-adaptativa baseada no conceito de expectativa de vida é aplicada para previsão de velocidade do vento e potência gerada em parques eólicos. O conceito de expectativa de vida é aplicado inicialmente no *particle swarm optimization* (PSO) utilizado para o treinamento de uma rede neural, e comparado com outras ferramentas para o treinamento da mesma rede. Em seguida é aplicado em máquinas *kernel* para definir qual máquina deve ser excluída de um conjunto (*challengers*), que realizam previsões de geração eólica de curtíssimo prazo em oito parques eólicos na Austrália. O método proposto é baseado em rastreamento competitivo, e o algoritmo lida com algumas dificuldades comuns do PSO e das máquinas *kernel*, como ótimos locais e o aumento do tamanho da matriz do *kernel* associado com o tempo e complexidades da memória. O método proposto considera sempre as novas informações recebidas pelo modelo, identificando assim mudanças nas séries temporais, evitando a perda abrupta de informações e mantendo um número controlado de exemplos uma vez que há uma seleção adaptativa das partículas e dos *kernels* ativos. Como resultado tem-se a redução da probabilidade de *overfitting* em ambas as aplicações, e no caso das máquinas *kernel*, o uso de dicionários menores quando comparados aos outros modelos. O uso do erro como indicador de performance apontou que a nova abordagem em ambas as aplicações (PSO e máquinas *kernel*) produz resultados melhores, além de reduções do custo computacional ao ser comparado com outros algoritmos como o *backpropagation*, *differential evolution* (DE), uma versão *online* do método de máquina de aprendizado extremo, persistência, e diferentes máquinas *kernel*.

Palavras-chave: Previsão, Velocidade do vento, Geração eólica, Parques eólicos *onshore*, Otimização por enxames de partículas, Máquinas *kernel*, Aprendizado de máquinas.

ABSTRACT

The increase of wind power generation has presented challenges to the electrical networks, requiring a decision-making tool capable of dealing with its intermittence. For this purpose, a self-adaptive approach based on the concept of life expectancy is introduced. The concept of life expectancy is applied to the PSO which is used for training a neural network and compared with other training tools. Then, it is adapted to kernel machines through an auto-adaptative multiple kernel learning algorithm, which is successfully used to produce very short-term wind power forecasts at eight wind farms in Australia. The proposed method is based on a competitive tracking method, and the algorithm deals with some common difficulties of PSO and kernel methods, e.g., local optima clustering and the increasing kernel matrix size associated with time and memory complexities and the overfitting problem. The proposed method always considers the new information received by the model, thus identifying changes in the time series, avoiding abrupt loss of information and maintaining a controlled number of examples, since there is an adaptive selection of particles and active kernels. As a result, reducing the probability of overfitting in both applications, and working with the smallest dictionary possible when compared to others kernel machines. The new method, compared to others, such as backpropagation, DE, an online version of the extreme learning machine, persistence and different kernel machines, obtains lowest errors and a reduction on execution time in general.

Keywords: Forecasting, Wind speed, Wind power, Wind farm, Particle swarm optimization, Kernel machines, Machine learning.

LISTA DE FIGURAS

Figura 1 – Comparação das abordagens (a) física e (b) estatística.	23
Figura 2 – Diagrama de blocos do sistema baseado na abordagem física.	35
Figura 3 – Curva de potência típica de um aerogerador.	43
Figura 4 – Previsões pontuais associadas a intervalos de confiança.	49
Figura 5 – FTDNN com arquitetura genérica.	57
Figura 6 – Diagrama de blocos dos processos do estudo de caso 1.	75
Figura 7 – Diagrama de blocos dos processos do estudo de caso 2.	75
Figura 8 – Diagrama de blocos dos processos do estudo de caso 3.	75
Figura 9 – Histograma do conjunto de dados WS entre 2009 e 2013.	76
Figura 10 – Diagrama de Taylor para 4 modelos (LGD).	80
Figura 11 – Diagrama de caixas do próximo passo (LGD).	81
Figura 12 – Performance do FTDNN sobre diferente números de geração máxima.	82
Figura 13 – Relação entre o tamanho do conjunto de treinamento, o MSE, e o tempo de execução.	84
Figura 14 – Relação entre o número de variáveis de entrada e o MSE.	86
Figura 15 – Relação entre o NNHL, LAG, e MSE.	87
Figura 16 – Relação entre o NNHL, LAG, e MSE.	88
Figura 17 – Melhores resultados para cada conjunto de variáveis de entrada.	89
Figura 18 – Performance da FTDNN para diferentes dados de entrada e algoritmos de treinamento usados para obter os valores de NNHL e LAG.	90
Figura 19 – Performance do PSO-AWF comparado aos métodos de aprendizado usando TE.	91
Figura 20 – Diagrama de blocos dos processos dos estudos de caso 4 e 5.	92
Figura 21 – Diagrama de blocos com todos os arranjos das máquinas KRLS aplicadas nos experimentos.	95
Figura 22 – Comparação entre o <i>approximate linear dependency</i> KRLS (ALD-KRLS) MIMO e ALD-KRLS M3C-KRLSW, $\nu = 1.00E-05$ e variando o número de LAG.	98
Figura 23 – Comparação entre o RMSPE versus quando iniciar uma nova máquina <i>kernel</i> , variando o tamanho máximo do dicionário.	99
Figura 24 – Comparação entre o RMSPE versus o tamanho do dicionário.	101
Figura 25 – Rosa dos ventos dos anos.	124

Figura 26 – Rosa dos ventos de janeiro.	125
Figura 27 – Rosa dos ventos de fevereiro.	126
Figura 28 – Rosa dos ventos de março.	127
Figura 29 – Rosa dos ventos de abril.	128
Figura 30 – Rosa dos ventos de maio.	129
Figura 31 – Rosa dos ventos de junho.	130
Figura 32 – Rosa dos ventos de julho.	131
Figura 33 – Rosa dos ventos de agosto.	132
Figura 34 – Rosa dos ventos de setembro.	133
Figura 35 – Rosa dos ventos de outubro.	134
Figura 36 – Rosa dos ventos de novembro.	135
Figura 37 – Rosa dos ventos de dezembro.	136

LISTA DE TABELAS

Tabela 1 – Horizonte de tempo para a previsão	28
Tabela 2 – Lista de <i>software</i> de previsão de energia eólica	42
Tabela 3 – Modelos MC-KRLS utilizados	72
Tabela 4 – Resultados da FTDNN com diferente valores de NNHL.	82
Tabela 5 – Resultados da FTDNN com diferente valores de NNHL (<i>5-fold</i>).	83
Tabela 6 – Resultados das FTDNN com diferentes tamanhos de conjunto de treinamento.	85
Tabela 7 – Informações sobre cada parque eólico.	93
Tabela 8 – Parâmetros usados pelos modelos.	95
Tabela 9 – RMSPE dos modelos de referência.	96
Tabela 10 – Coeficiente de correlação de pearson	96
Tabela 11 – RMSPE de todos os modelos	97
Tabela 12 – Tempo de execução (s)	101

LISTA DE ABREVIATURAS E SIGLAS

AE	<i>absolute error</i>
AEC	<i>adaptive exponential combination</i>
AF	<i>aging factor</i>
ALD	<i>approximate linear dependency</i>
ALD-KRLS	<i>approximate linear dependency KRLS</i>
ANN	<i>artificial neural network</i>
AR	<i>autoregressive</i>
ARCH	<i>autoregressive conditional heteroscedastic</i>
ARIMA	<i>autoregressive integrated moving average</i>
ARMA	<i>autoregressive moving average</i>
ARMAX	<i>autoregressive moving average with exogenous input</i>
ARX	<i>autoregressive with exogenous input</i>
AWPPS	<i>armines wind power prediction system</i>
BJ	<i>Box-Jenkins gas furnace data</i>
BP	<i>backpropagation</i>
CFD	<i>computational fluid dynamics</i>
CMPSODMO	<i>cooperative multi-swarm PSO for dynamic multi-objective optimizations</i>
DAL	<i>distance active learning</i>
DE	<i>differential evolution</i>
DMS-PSO	<i>dynamic multiswarm PSO</i>
DTU	<i>Danmarks Tekniske Universitet</i>
ELM	<i>extreme learning machines</i>
EPSO	<i>enhanced particle swarm optimization</i>
FAR	<i>false alarm rate</i>
FTDNN	<i>focused time delay neural network</i>
GFNN	<i>generalized feedforward neural network</i>
HIRLAM	<i>high resolution limited area model</i>
HPC	<i>high performance computer</i>
IC	<i>inteligência computacional</i>
IQR	<i>interquartile range</i>

ITSM	<i>improved time series method</i>
KRLS	<i>kernel recursive least-squares</i>
KRLS-T	<i>KRLS tracker</i>
KRR	<i>kernel ridge regression</i>
LE	<i>life expectancy</i>
LGD	<i>Data set A: laser generated data</i>
LQR	<i>local quantile regression</i>
LS	<i>least squares</i>
LS-SVM	<i>least squares SVM</i>
MA	<i>moving average</i>
MAE	<i>mean absolute error</i>
MAPE	<i>mean absolute percentage error</i>
MAS	<i>multiple architecture system</i>
MC	<i>multiple challengers</i>
MC-KRLS	<i>multiple challengers kernel recursive least-squares</i>
MCP	<i>measure-correlate-predict</i>
ME	<i>mean error</i>
MG	<i>Mackey-Glass chaotic time series</i>
MHNN	<i>modified hybrid neural network</i>
MI	<i>mutual information</i>
MIMO	<i>multiple input and multiple output</i>
MKL	<i>multiple kernel learning</i>
MLP	<i>multi-layer perceptrons</i>
MLR	<i>multiple linear regression</i>
MOS	<i>model output statistics</i>
MRI	<i>meteo-risk index</i>
MSE	<i>mean square error</i>
NPRI	<i>normalized prediction risk index</i>
NWP	<i>numerical weather prediction</i>
NYISO	<i>New York independent system operator</i>
NYSERDA	<i>New York state energy research and development authority</i>
OS-ELM	<i>online sequential extreme learning machine</i>

PAL	<i>pool of regressors active learning</i>
PCLPSO	<i>parallel comprehensive learning PSO</i>
PDF	<i>probability density function</i>
PoD	<i>probability of detection</i>
PSO	<i>particle swarm optimization</i>
PSO-AWF	<i>particle swarm optimization with aging and weakening factor</i>
RBF	<i>radial basis function</i>
RLS	<i>recursive least-squares</i>
RMSD	<i>root mean square deviation</i>
RMSE	<i>root mean square error</i>
RMSPE	<i>root mean square percentage error</i>
RSAL	<i>residual regression active learning</i>
SARIMA	<i>seasonal autoregressive integrated moving average</i>
SCADA	<i>supervisory control and data acquisition</i>
SISO	<i>single input and single output</i>
SVM	<i>support vector machines</i>
SW-KRLS	<i>sliding-window KRLS</i>
TC1	Estudo de caso 1, <i>test case 1</i>
TC2	Estudo de caso 2, <i>test case 2</i>
TC3	Estudo de caso 3, <i>test case 3</i>
TC4	estudo de caso 4, <i>test case 4</i>
TC5	estudo de caso 5, <i>test case 5</i>
TE	tentativa e erro
WAsP	<i>wind atlas analysis and application program</i>
WF	<i>weakening factor</i>
WGP	<i>warped gaussian process</i>
WS	<i>wind dataset</i>

LISTA DE SÍMBOLOS

LE	Expectativa de vida, adimensional
AF	<i>Aging factor</i> utilizado para reproduzir o envelhecimento do indivíduo, adimensional
WF	<i>Weakening factor</i> utilizado para reproduzir escolhas que acelerem o envelhecimento do indivíduo, adimensional
t	Instante de tempo, minutos
Δ	Resolução de tempo da previsão
$v_{t+\Delta t}$	Velocidade do vento média entre o instante t e $t + \Delta$, metros por segundo
$P_{t+\Delta t}$	Potência média entre o instante t e $t + \Delta$, Watts
T	Horizonte de tempo que indica o tempo total do período a ser previsto, minutos
$\hat{P}_{t+\Delta t}$	Potência esperada entre o t e $t + \Delta$, Watts. Note que variáveis que usam o sinal de circunflexo indicam variáveis prevista e não medidas
$\hat{Y}_{t+\Delta t}$	Saída esperada entre o instante t e $t + \Delta$
Y_t	Saída medida no instante t
P_t	Potência medida no instante t
$\hat{u}_{t+\Delta}$	Velocidade do vento prevista pelo NWP no instante $t + \Delta$, metros por segundo
$\hat{\Theta}_{t+\Delta}$	Direção do vento prevista pelo NWP no instante $t + \Delta$, graus
$\hat{x}_{t+\Delta}$	Padrões diversos previstos pelo NWP no instante $t + \Delta$, graus
$f(\cdot)$	Função f
v_0	Velocidade do vento mínima para que a turbina eólica inicie seu funcionamento, metros por segundo
v_n	Velocidade do vento mínima para que a turbina eólica inicie a entregar potência convertida, metros por segundo
v_{max}	Velocidade do vento máxima para que a turbina eólica funcione, metros por segundo
P	Potência calculada da turbina eólica, Watts

ρ_{ar}	Densidade do ar, quilograma por metro cúbico
C_p	Coefficiente de performance da turbina, adimensional Watts
A_r	Área varrida pelo rotor da turbina, metros quadrados
v	Velocidade do vento na altura do <i>hub</i> , metros por segundo segundo
$e_{t+\Delta}$	Erro da previsão
N	Número total de pontos da série de dados
σ	Desvio padrão
r	Coefficiente de correlação cruzada
(x_i, y_i)	Pares de dados históricos (medições), onde x indica variável de entrada do modelo e y a variável de saída
LAG	Linha de atrasos, número de pontos de uma determinada variável utilizada
F	Parâmetro controle de mutação do algoritmo DE
CR	Parâmetro controle de cruzamento do algoritmo DE
NP	Parâmetro tamanho da população do algoritmo DE
$v_i(t+1)$	Nova velocidade da partícula i do algoritmo PSO
$v_i(t)$	Velocidade da partícula i do algoritmo PSO
C_1	Constante social do algoritmo PSO
R_1	Número randômico entre 0 e 1 do algoritmo PSO
C_2	Constante cognitiva do algoritmo PSO
R_2	Número randômico entre 0 e 1 do algoritmo PSO
$x_i(t)$	Posição atual da partícula i do algoritmo PSO
$x_i(t+1)$	Nova posição da partícula i do algoritmo PSO
ω	Peso de inércia do algoritmo PSO
PA	Idade da partícula, adimensional
η_m	Peso para o kernel a ser combinado
k_m	Máquina <i>kernel</i> de índice m
$M2C$	Modelo com duas máquinas <i>kernel</i>
$M3C$	Modelo com três máquinas <i>kernel</i>

<i>NNHL</i>	<i>Number of neurons in the hidden layer</i>
<i>D</i>	Direção do vento
<i>p</i>	Pressão
<i>T</i>	Temperatura
<i>v</i>	<i>Approximate linear dependency treshold</i> do algoritmo ALD-KRLS
<i>M</i>	Tamanho do dicionário para os algoritmos ALD-KRLS, SW-KRLS e KRLS-T
σ	Largura do <i>kernel</i> gaussiano para os algoritmos ALD-KRLS, SW-KRLS e KRLS-T
<i>c</i>	Parâmetro de regularização do algoritmo SW-KRLS
λ	Fator de esquecimento do algoritmo KRLS-T
<i>jitter</i>	Ruído utilizado para evitar o erro de arredondamento do algoritmo KRLS-T
<i>sn2</i>	Relação entre o ruído e o sinal do algoritmo KRLS-T
<i>NOHN</i>	<i>Number of hidden nodes</i> , NNHL é uma adaptação do termo do algoritmo OS-ELM

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Motivação	22
1.2	Objetivo	24
1.3	Contribuições da tese	25
1.4	Produção científica	25
1.5	Organização do trabalho	25
2	REFERENCIAL TEÓRICO	27
2.1	Definições da previsão eólica	27
2.2	Horizontes de tempo	27
2.3	Visão geral do estado da arte dos modelos de previsão eólica	28
2.3.1	<i>Modelo de referência</i>	30
2.3.2	<i>Modelos estatísticos</i>	30
2.3.2.1	<i>Abordagem estatística convencional</i>	31
2.3.2.2	<i>Abordagem estatística não convencional</i>	32
2.3.3	<i>Modelos físicos</i>	34
2.3.4	<i>Modelos híbridos</i>	36
2.3.5	<i>Modelos de previsão com correlação espacial</i>	38
2.3.6	<i>Modelos de previsão regionais (Upscaling)</i>	39
2.3.7	<i>Modelos de previsão probabilísticos</i>	39
2.3.7.1	<i>Modelos de previsão probabilísticos paramétricos</i>	39
2.3.7.2	<i>Modelos de previsão probabilísticos não-paramétricos</i>	40
2.3.8	<i>Modelos de previsão marítimos</i>	41
2.3.9	<i>Ferramentas de previsão disponíveis</i>	42
2.4	Curva de potência de aerogeradores	43
2.5	Avaliação de performance dos modelos	44
2.5.1	<i>Decomposição do RMSE da previsão da velocidade do vento</i>	46
2.5.2	<i>Avaliação da performance dos modelos probabilísticos</i>	47
2.6	Incerteza das previsões de geração eólica	48
2.6.1	<i>Índices de risco da previsão</i>	50
2.7	Valor da previsão da geração eólica	52

2.8	Conclusões	53
3	DESCRIÇÃO DO PROBLEMA E FERRAMENTAS DE INTELIGÊN- CIA COMPUTACIONAL	55
3.1	Ferramentas para o método <i>offline</i>	56
3.1.1	<i>Particle swarm optimization</i>	58
3.1.2	<i>Trabalhos relacionados</i>	58
3.2	<i>Particle swarm optimization with aging and weakening factors (PSO-AWF)</i>	60
3.3	Ferramentas para o método <i>online</i>	62
3.3.1	<i>Máquinas kernel</i>	63
3.3.2	<i>Trabalhos relacionados</i>	64
3.4	<i>Multiple challengers kernels recursive least-squares (MC-KRLS)</i>	70
3.4.1	<i>Quando criar uma nova matriz do kernel</i>	70
3.4.2	<i>Como calcular a previsão combinada (\hat{Y})</i>	71
3.4.3	<i>Quando apagar a matriz de um kernel</i>	71
3.5	Conclusões	73
4	ESTUDO DE CASO	74
4.1	Método <i>offline</i>	74
4.1.1	<i>Conjuntos de dados</i>	74
4.1.2	<i>Definições de parâmetros</i>	77
4.1.3	<i>Resultados</i>	78
4.1.3.1	<i>Estudo de caso 1</i>	79
4.1.3.2	<i>Estudo de caso 2</i>	83
4.1.3.3	<i>Estudo de caso 3</i>	87
4.2	Método <i>online</i>	91
4.2.1	<i>Conjunto de dados</i>	93
4.2.2	<i>Definições de parâmetros</i>	94
4.2.3	<i>Resultados</i>	94
4.2.3.1	<i>Estudo de caso 4</i>	95
4.2.3.2	<i>Estudo de caso 5</i>	97
4.3	Conclusões	102
5	CONSIDERAÇÕES FINAIS	103
5.1	Pesquisas futuras	105

REFERÊNCIAS	106
APÊNDICES	123
APÊNDICE A – ROSA DOS VENTOS DO CONJUNTO DE DADOS WS	123

1 INTRODUÇÃO

Fenômenos não lineares são encontrados em muitos problemas de engenharia. Tradicionalmente as técnicas de processamento de sinais são lineares, as quais não são aptas a extrair os padrões complexos e não lineares que possam estar presentes em fenômenos não lineares. Portanto, problemas envolvendo análise de dados não lineares têm sido tradicionalmente abordados por filtros polinomiais (MATHEWS; SICURANZA, 2000), os quais são extensões diretas de muitos métodos lineares, ou através de redes neurais artificiais, do inglês *artificial neural network* (ANN) (HAYKIN, 2007), que são aptas a aprender essas características não lineares.

Em Haykin (2007), aprendizado é definido como o ‘processo pelo qual os parâmetros livres de uma ANN são adaptados através das estimulações oferecidas pelo seu ambiente’. Conseqüentemente, o objetivo do processo de aprendizado é fazer com que a ANN responda da maneira desejada ao seu ambiente. O processo de aprendizado em si, garante que a performance melhore com a sua experiência.

Dois conceitos importantes são a capacidade de aprendizado e generalização (COR- TES; VAPNIK, 1995). A capacidade de aprendizado se refere ao quão capaz é uma máquina em representar funções complexas e não lineares. A capacidade de generalização representa o quão bem uma máquina representa novos dados. Claramente existe uma proporcionalidade inversa entre estas capacidades, enquanto que uma alta capacidade de aprendizado permitirá que os padrões dos dados de treinamento sejam representados precisamente, a capacidade de generalização será pequena, levando a uma baixa performance ao serem apresentados novos dados.

Ao contrário das técnicas lineares, máquinas de aprendizado não lineares como as ANN requerem mais operações, e normalmente envolvem problemas de otimização não lineares com múltiplos mínimos ou máximos locais. Uma alternativa a isso são as máquinas *kernel* (SCHÖLKOPF; SMOLA, 2002; SHAWE-TAYLOR; CRISTIANINI, 2004). Métodos *kernel* são técnicas de aprendizado de máquina que utilizam uma arquitetura menos complexa e proporciona uma abordagem direta ao transformar problemas não lineares em problemas de otimização convexa. Máquinas de aprendizado são utilizadas para classificação, regressão e agrupamento.

Nesta tese é estudada a aplicação de ANN e máquinas *kernel* ao problema de previsão de velocidade do vento e potência gerada em usinas de conversão eólico elétricas normalmente nominados parques eólicos.

1.1 Motivação

A geração eólica continua crescendo em todo mundo (LEE; ZHAO, 2020). Neste contexto, a exposição da produção à intermitência característica do vento e o controle restrito das turbinas eólicas, justificam o desenvolvimento de novos modelos de previsão de geração eólica (ARAGHI *et al.*, 2020). Este é um ponto chave para assegurar a integração dos parques eólicos ao sistema elétrico de potência (SSEKULIMA *et al.*, 2016). Assim, a operação confiável e econômica de sistemas de energia com alta penetração de energia eólica exige a utilização de modelos de previsão precisos (VARGAS *et al.*, 2019).

A previsão da energia eólica tem horizontes de tempo diferentes de acordo com sua aplicação (ROUNGKVIST; ENEVOLDSEN, 2020). A classificação na escala de tempo não é um tópico de consenso geral, mas pode ser classificada em curtíssimo prazo (alguns segundos à 30 minutos à frente), curto prazo (30 minutos a 24 horas à frente), médio prazo (24 horas a uma semana adiante), e de longo prazo (uma semana a anos à frente) (SOMAN *et al.*, 2010; SFETSOS, 2011; ZHAO *et al.*, 2011; ROUNGKVIST; ENEVOLDSEN, 2020).

Uma maneira de classificar os modelos de previsão da velocidade do vento e/ou potência gerada é se envolvem ou não o uso do *numerical weather prediction* (NWP), e modelos de séries temporais. Normalmente, os modelos de previsão que usam previsões NWP superam as abordagens de séries temporais em horizontes de tempo acima de 6 horas à frente. Se o horizonte de previsão não for muito longo, a velocidade e a potência do vento podem ser previstas apenas usando métodos de análise de séries temporais, sem recorrer a previsões meteorológicas reais. Modelos de séries temporais diretas¹ são aqueles que usam valores observados recentes de vento e outras variáveis para prever a velocidade futura do vento (GIEBEL; KARINIOTAKIS, 2017). Historicamente os modelos são categorizados como físicos ou estatísticos, e com o avanço da área foi criado um consenso que normalmente métodos ótimos utilizam as duas categorias em conjunto, conhecidos como modelos híbridos (GIEBEL; KARINIOTAKIS, 2017).

Os modelos físicos usam descrições de baixa atmosfera (troposfera), características geográficas e de obstáculos para prever o fluxo do vento, podendo utilizar modelos de previsão numérica do tempo, do inglês NWP. Os modelos NWP preveem variáveis meteorológicas como velocidade do vento, direção do vento, pressão entre outras, usando informações espaciais e temporais tridimensionais com base em modelos de dinâmica de fluidos computacional, do inglês *computational fluid dynamics* (CFD). As previsões de energia eólica podem ser obtidas

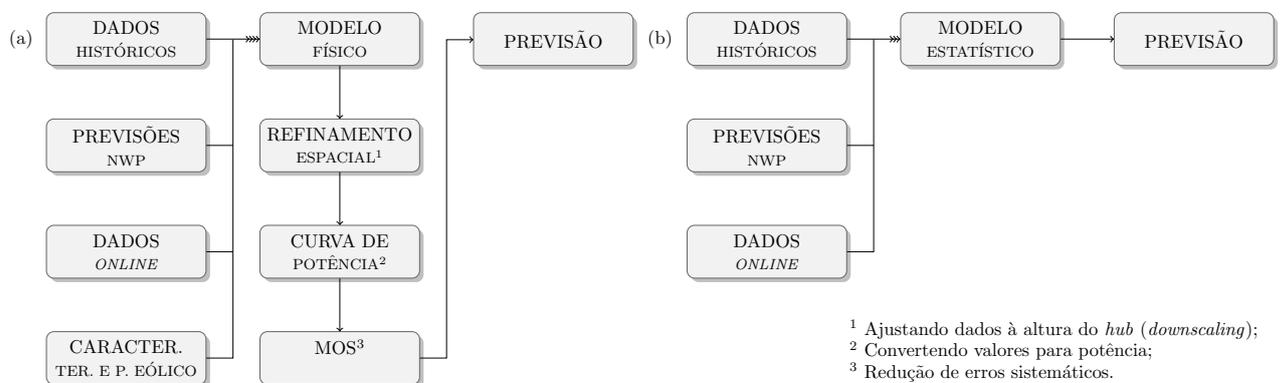
¹ O termo direto indica a não conversão da variável prevista em potência.

com base no desempenho de uma turbina eólica ou de um parque eólico usando os resultados do NWP (ANDRADE; BESSA, 2017; ZHANG *et al.*, 2017). Por sua complexidade e um tempo de desenvolvimento considerável, modelos físicos têm uso comum em previsões acima de 6 horas (FOLEY *et al.*, 2012; SANTHOSH *et al.*, 2020); outro ponto importante é a necessidade de condições iniciais precisas de parques eólicos que nem sempre podem ser garantidas (LIU; CHEN, 2019). No entanto, modelos físicos podem representar fenômenos climáticos, como as frentes climáticas (frente frias e quentes) e tempestades.

Os modelos estatísticos e de aprendizado de máquina procuram uma relação entre variáveis de entrada e saída que represente o comportamento oriundo de séries temporais com base em dados históricos e/ou do NWP (GIEBEL; KARINIOTAKIS, 2017). O tempo de desenvolvimento destes modelos são menores do que os modelos físicos. Como exemplos têm-se o modelo autorregressivo de médias móveis com entrada exógena, do inglês *autoregressive moving average with exogenous input* (ARMAX) (LYDIA *et al.*, 2016), o modelo autorregressivo integrado de médias móveis, do inglês *autoregressive integrated moving average* (ARIMA) (SIM *et al.*, 2019), ANN (WANG *et al.*, 2019), as máquinas de vetor de suporte, do inglês *support vector machines* (SVM) (PAWAR *et al.*, 2019), a lógica *fuzzy* (KHORRAMDEL *et al.*, 2018; LIU *et al.*, 2019), e as máquinas de aprendizado extremo, do inglês *extreme learning machines* (ELM) (LI *et al.*, 2019; ACIKGOZ *et al.*, 2020).

A Figura 1 ilustra a comparação das abordagens física e estatística através de diagramas de blocos que representam etapas do processo de previsão de maneira simplificada. O modelo será discutido com maiores detalhes na Seção 2.3.3.

Figura 1 – Comparação das abordagens (a) física e (b) estatística.



Fonte: Elaborado pelo autor (2021).

Por fim os modelos híbridos, com o objetivo de melhorar o desempenho da previsão, combinam diferentes metodologias para tirar proveito de cada uma (QIAN *et al.*, 2019), como

modelos baseados em: ponderação, técnicas de pré-processamento de dados, seleção de parâmetros e técnicas de otimização, e técnicas de processamento de erro (TIAN *et al.*, 2018; MISHRA; DASH, 2019; QU *et al.*, 2019; SINGH *et al.*, 2019; TIAN *et al.*, 2020). Entre eles, abordagens baseadas em métodos de decomposição de séries temporais têm sido frequentemente relatadas, em que a série temporal original pode ser decomposta em diferentes subséries e modelada de forma mais eficaz do que a série temporal original (LIU *et al.*, 2020).

1.2 Objetivo

Os métodos estatísticos usados para a previsão da velocidade e potência gerada em parques eólicos são geralmente lineares, apesar da natureza não linear do vento e são normalmente empregados em locais únicos. Considerando este aspecto fundamental de não linearidade do problema, o presente estudo tem como objetivo geral investigar e propor modificações em algoritmos de aprendizagem através de uma releitura do conceito de expectativa de vida, do inglês *life expectancy* (LE). Para isto, foi focado o efeito da LE na diversidade cultural de uma colônia animal. A LE é uma medida estatística do quanto um organismo pode viver, e em uma certa idade é esperado que a vida cesse. A LE é baseada em diversas características, como por exemplo, o ano de nascimento, idade atual, e outros fatores demográficos. Primeiro a investigação é feita no treinamento de redes neurais focadas em atrasos temporais, do inglês *focused time delay neural network* (FTDNN), em seguida aplicado aos métodos *kernel*, que podem fornecer processamento linear de recursos não lineares para um e vários locais. Máquinas *kernel* retêm as propriedades do processamento linear, como algoritmos de aprendizado rápido e uma solução ótima única, ao mesmo tempo que torna possível capturar algumas não linearidades. As máquinas *kernel* combinam a teoria de aprendizagem estatística para otimizar a generalização (VAPNIK, 2006), com programação matemática para encontrar soluções de forma eficiente, bem como para melhorar a medida de similaridade entre pontos para lidar com problemas de não linearidade. Como objetivos específicos têm-se:

- Propor um algoritmo de aprendizado de ANN baseado em uma versão modificada do PSO, para encontrar erros menores ou similares com redução significativa do custo computacional, quando comparado às ferramentas de referência;
- Propor um algoritmo de aprendizado de múltiplas máquinas *kernel*, para encontrar erros menores ou similares com dicionários (matrizes *kernel*) menores e não estáticos, quando comparados às estimativas realizadas por suas respectivas máquinas *kernel* com dicionários

únicos;

- Aplicar os algoritmos propostos na previsão da velocidade do vento em um parque eólico brasileiro, e geração de um conjunto de parques australianos.

1.3 Contribuições da tese

As principais contribuições desta tese estão relacionadas a seguir:

- Propor uma versão simplificada do conceito LE, baseado em duas características, chamadas fator de envelhecimento, do inglês *aging factor* (AF), e fator de enfraquecimento, do inglês *weakening factor* (WF);
- Propor um método simples para realizar um controle adaptativo das partículas do PSO utilizando o AF e WF (PSO-AWF) e avaliar o comportamento de busca do novo controle aplicado ao PSO;
- Propor um esquema de aprendizado de múltiplos *kernels* chamado aqui de múltiplos desafiantes, do inglês *multiple challengers* (MC), com controle adaptativo do número de *kernels* utilizados pelo preditor baseado na versão simplificada do LE, e avaliar o comportamento de busca do novo controle aplicado ao aprendizado de múltiplos *kernels*.

1.4 Produção científica

Ao longo do desenvolvimento desta tese os seguintes artigos científicos foram produzidos:

- Bezerra, E.C., Leão, R.P.S. and Braga, A.P.d.S. A Self-Adaptive Approach for Particle Swarm Optimization Applied to Wind Speed Forecasting. *J Control Autom Electr Syst* 28, 785–795 (2017). <https://doi.org/10.1007/s40313-017-0339-6>;
- Bezerra, E.C., Pinson, P., Leão, R.P.S. and Braga, A.P.d.S. A Self-Adaptive Multikernel Machine Based on Recursive Least-Squares Applied to Very Short-Term Wind Power Forecasting. *IEEE Access* 9, 104761-104772 (2021). <https://doi.org/10.1109/ACCESS.2021.3099999>.

1.5 Organização do trabalho

O restante deste trabalho está organizado da seguinte forma, o Capítulo 2 apresenta uma visão geral dos modelos de previsão, métodos de aferição de erros, e o valor econômico da previsão. O objetivo do capítulo é introduzir o leitor a diferentes aspectos que constituem a

previsão e ilustrar as diferentes linhas de pesquisa que têm sido exploradas até a presente data. No Capítulo 3 são apresentadas a formulação do problema, e de forma sucinta, as ferramentas de inteligência computacional (IC) que são utilizadas nesta tese. O Capítulo 4 se detém à apresentação de estudos de casos para validar os modelos propostos através da análise dos resultados obtidos. O desempenho dos modelos *offline* são medidos em diferentes horizontes de previsão (5', 30', 1h, 3h e 6h), enquanto o modelo *online* foca no horizonte de 5 minutos. No Capítulo 5 são apresentadas as conclusões da tese, e também apontadas pesquisas futuras.

2 REFERENCIAL TEÓRICO SOBRE PREVISÃO DE GERAÇÃO EÓLICA

Neste capítulo é apresentada uma visão geral de conceitos, modelos de previsão, métodos de aferição de erros, e valor econômico da previsão encontrados na literatura no campo da previsão da geração eólica. Seu objetivo é apresentar o estado-da-arte sobre os diferentes aspectos que constituem a previsão da geração eólica e as diferentes linhas de pesquisa exploradas, a partir da revisão bibliográfica.

2.1 Definições da previsão eólica

A previsão eólica, quer seja de velocidade do vento ou de potência gerada, no instante t , para um intervalo no tempo de $t + \Delta$, é a média da velocidade do vento $v_{t+\Delta|t}$ ou potência $P_{t+\Delta|t}$, que é esperada para um parque eólico durante período de tempo Δ . Previsões são realizadas para um horizonte de tempo T , indicando o tempo total do período a ser previsto, por exemplo 24 ou 72 horas a frente, no futuro. O intervalo de tempo Δ denota a resolução de tempo da previsão. O comprimento de Δ está relacionado ao de T .

Normalmente, para horizontes da ordem de 24 a 72 horas, Δ é de hora em hora. Neste caso, as variações de potência ocorridas nestes intervalos de 1 hora, por exemplo, e seus impactos não são consideradas. Esta convenção também vem do fato de que as informações oriundas dos modelos NWP utilizadas na previsão da velocidade do vento, são dadas como valores constantes para o passo à frente.

Na prática, o valor da medida da potência $P_{t+\Delta}$ é derivada a partir da média de medições (por exemplo, de 15 min.), que podem ser valores da potência instantânea ou valores de energia integradas, dependendo do sistema de aquisição de dados.

É importante ressaltar que $\hat{P}_{t+\Delta|t}$ é chamado de previsão pontual, porque é apenas um único valor. Existem outros tipos de previsões, como as previsões probabilísticas, que geram uma distribuição de probabilidade da previsão para cada intervalo de tempo feito, e podem ser representados através de densidade ou percentuais da previsão (Seção 2.3.7).

2.2 Horizontes de tempo

Um sistema de previsão é caracterizado pelo seu horizonte de tempo T , que é o período de tempo futuro pelo qual a produção de energia eólica será prevista (por exemplo, no dia seguinte), como foi explicado na seção anterior.

Na previsão de geração eólica, as fronteiras de tempo que separam os diferentes horizontes temporais não são definidas unanimemente. Diferentes autores têm proposto diferentes fronteiras para cada categoria de horizonte de tempo, e normalmente seus limites são definidos pela sua aplicação, que são inúmeras (CUTLER, 2006). Definir o objetivo da previsão da geração eólica é o primeiro passo para elaborar o projeto de um sistema de previsão. Para definir esse objetivo é necessário pelo menos identificar o horizonte entre o momento atual e a previsão (por exemplo, dias, semanas, meses), e a frequência das previsões (horária). A Tabela 1 apresenta aplicações da previsão de acordo com o horizonte de tempo.

Tabela 1 – Horizonte de tempo para a previsão

Horizonte de tempo	Aplicação
Segundos à 1 hora	Controle ativo do aerogerador Balanço carga-frequência
1 à 6 horas	Operação imediata dos aerogeradores Planejamento econômico de despacho de carga
6 horas até 3 dias	Manutenção e operação Despacho da geração Segurança operacional do mercado de energia Mercado de energia de compensação Programação e despacho da geração Mercado de energia
3 dias até anos	Planejamento de projetos e orçamentos
Anos até décadas	Riscos volumétricos

Fonte: Adaptado de Soman *et al.* (2010), Chang (2014), e Rounkvist e Enevoldsen (2020)

De maneira geral são utilizados quatro horizontes diferentes: curtíssimo prazo (poucos segundos até 30 min), curto prazo (30 min até 24 h), médio prazo (24 h até 1 semana), e longo prazo (de 1 semana até anos) (SOMAN *et al.*, 2010; SFETSOS, 2011; ZHAO *et al.*, 2011; ROUNGKVIST; ENEVOLDSEN, 2020). Este trabalho se baseia nesta divisão, e foca nas previsões de curtíssimo e de curto prazo. Quando definido o objetivo da previsão, então pode ser escolhido o modelo a ser utilizado.

2.3 Visão geral do estado da arte dos modelos de previsão eólica

O uso clássico de séries temporais para previsão se baseia na análise de séries históricas da variável a ser prevista e de outras variáveis que contribuem para a previsão. Estes modelos são muitas vezes capazes de oferecer boas previsões. Entretanto, a natureza caótica e não estacionária da atmosfera não permite computar previsões de séries temporais para a

geração eólica para horizontes maiores que 4 a 6 horas (MILLIGAN *et al.*, 2004). Previsões para horizontes mais distantes em geral fazem uso de valores previstos oriundos do NWP, como uma das entradas do modelo de previsão que pode ser um modelo estatístico, físico ou híbrido. Entretanto, alguns problemas surgem devido à natureza do modelo NWP e à dificuldade de determinar uma curva de potência confiável para a turbina e parque eólico.

Modelos NWP (ZHANG *et al.*, 2017; ANDRADE; BESSA, 2017) são baseados em resolução espacial com equações que descrevem a atmosfera. Devido a restrições computacionais (FOLEY *et al.*, 2012), essas equações são resolvidas em grades de três dimensões. As grades variam da ordem de dezenas a centenas de quilômetros, de acordo com o modelo NWP, com diferentes níveis verticais e diferentes números de nós no plano horizontal. Esta característica dos modelos NWP levam a duas situações quando utilizados para a previsão de geração eólica. A primeira é que o parque é localizado em um dos nós da grade. Mesmo que as variáveis meteorológicas previstas correspondam à localização geográfica próxima ao parque, não é necessariamente as correspondentes ao local do parque. Em alguns casos os valores são interpolados para o local do parque. A segunda, é que a grade usada no modelo NWP não consegue enxergar características menores que a malha da grade¹. Dependendo da forma da curva de potência, estes erros podem ser ampliados. Geralmente se admite que uma grande porção do erro da previsão de geração eólica é oriunda de erros do modelo NWP (KARINIOTAKIS *et al.*, 2004).

A previsão da geração eólica pode ser dividida em duas etapas. A primeira, conhecida como *downscaling* que será detalhada mais à frente, e a segunda é a modelagem da curva de potência, correspondente à transformação do vento em potência escalonado à altura do cubo (*hub*) da turbina eólica.

As abordagens de previsões de potência eólica normalmente envolvem o modelo NWP e eventualmente medições diversas (pressão, umidade do ar, temperatura, etc.) como entrada de dados. Esses métodos que utilizam os dados provenientes de um modelo NWP e medições são capazes de fornecer uma melhor previsão para as próximas 24-48 horas. Alternativamente, modelos podem ser construídos recebendo apenas medições como entrada (potência de geração, velocidade, etc.). Entretanto, o desempenho dos métodos baseados somente em medições pode ser aceitável para uma janela até 6 horas à frente. Para horizontes maiores, é necessária a inclusão dos dados do NWP. A inclusão de medições como entrada junto com o NWP contribui de maneira positiva para o desempenho da primeira faixa de previsão (0-6 horas).

¹ O tamanho da malha da grade varia entre modelos NWP diferentes.

Modelos utilizando apenas o NWP não superam o modelo de Persistência, descrito na Seção 2.3.1, nestas primeiras horas (KARINIOTAKIS *et al.*, 2004; GIEBEL; KARINIOTAKIS, 2017).

De maneira geral, os modelos podem ser classificados como os que utilizam os dados de NWP como entrada ou não. Em relação aos métodos que incorporam os dados de NWP, existem duas linhas principais: modelos estatísticos e físicos. Em alguns modelos a combinação dos dois é utilizada, sendo verdade que ambas as abordagens possam ser necessárias para obter boas previsões (SIEBERT, 2008; GIEBEL; KARINIOTAKIS, 2017).

O uso clássico de séries temporais para previsão se baseia na análise de valores históricos da variável a ser prevista e de variáveis outras que contribuem para a previsão. Estes modelos são muitas vezes capazes de oferecer boas previsões. Revisões do estado da arte sobre previsões de geração eólica estão disponíveis em Giebel *et al.* (2011), Okumus e Dinler (2016), Qian *et al.* (2019), Messner *et al.* (2020).

2.3.1 Modelo de referência

É interessante utilizar operacionalmente ferramentas avançadas de previsão do vento apenas se esta ferramenta for capaz de superar os resultados de técnicas simples resultantes de considerações sem esforços especiais de modelagem. Essas técnicas simples são usadas como referência para avaliar as ferramentas avançadas. A referência mais comum é o modelo de persistência que define a velocidade do vento futura como o último valor de velocidade estimado ou medido. Enquanto sua implementação é trivial, sua performance é considerada aceitável para previsões de curtíssimo e curto prazos. A previsão do intervalo de tempo Δ (i.e., segundos, minutos, horas,...) a frente é dada por $\hat{Y}_{t+\Delta} = Y_t$, onde $\hat{Y}_{t+\Delta}$ é o valor previsto do vento em um instante $t + \Delta$, Y_t é o valor da velocidade do vento no instante t .

2.3.2 Modelos estatísticos

Modelos estatísticos na sua forma pura tentam encontrar relações entre as variáveis incluindo os dados do NWP e dados de medições *online* (principalmente potência, mas também velocidade e direção do vento, se disponíveis). Normalmente são empregadas técnicas recursivas para encontrar os valores de potência relacionado aos dados de entrada escolhidos para aquele modelo (KARINIOTAKIS *et al.*, 2004; GIEBEL; KARINIOTAKIS, 2017).

Frequentemente, modelos caixa-preta como ANN são usados. Algumas abordagens exploram o conhecimento sobre as propriedades da potência eólica para definir a estrutura do

modelo (modelos caixa-cinza). Alguns modelos podem ser expressos analiticamente; outros, como ANN, não podem. Os modelos estatísticos também podem ser usados para fornecer previsão da velocidade do vento. Entretanto, é normalmente negligenciado e um modelo único é desenvolvido fornecendo diretamente a potência gerada (KARINIOTAKIS *et al.*, 2004; GIEBEL; KARINIOTAKIS, 2017).

A seguir, um exemplo simplificado é dado de como um modelo estatístico pode ser formulado. O modelo usa o NWP e produções medidas (se disponível *online*) para prever potência gerada. A forma geral do modelo é $\hat{P}_{t+\Delta} = f(P_t, \hat{u}_{t+\Delta}, \hat{\theta}_{t+\Delta}, \hat{x}_{t+\Delta})$, onde $\hat{P}_{t+\Delta}$ é a potência prevista para o tempo $t + \Delta$ realizado no tempo t , P_t são medições de produções passadas no tempo t , $\hat{u}_{t+\Delta}$ é a previsão da velocidade do vento fornecida pelo NWP no tempo $t + \Delta$, $\hat{\theta}_{t+\Delta}$ é a previsão da direção do vento fornecida pelo NWP no tempo $t + \Delta$, e $\hat{x}_{t+\Delta}$ são padrões para outras variáveis disponíveis previstas pelo NWP para o tempo $t + \Delta$.

A função $f(\cdot)$ pode ser por exemplo uma ANN ou uma rede *fuzzy*-neural, entre outras. Previsões com passos múltiplos podem ser geradas tanto pelo desenvolvimento de um modelo específico para cada horizonte ou utilizando um modelo de uma maneira mais interativa. Por exemplo, para produzir previsões para $t + 2$, a previsão $\hat{P}_{(t+1)}$ para $t + 1$ é uma entrada do modelo no lugar da potência medida (KARINIOTAKIS *et al.*, 2004).

A modelagem da curva de potência e das especificidades do local podem ser substituídas por um único passo que converte diretamente as variáveis de entrada (NWP, dados *online*) para potência.

Muitas abordagens são utilizadas ao ser realizada a previsão através de modelos estatísticos (LEI *et al.*, 2009; FOLEY *et al.*, 2010; SOMAN *et al.*, 2010; BHASKAR *et al.*, 2010; DONGMEI *et al.*, 2011; FOLEY *et al.*, 2012; KARAKUŞ *et al.*, 2017). Algumas destas abordagens serão descritas de maneira resumida nesta seção.

2.3.2.1 Abordagem estatística convencional

Modelos estatísticos auto-regressivos como, do inglês *autoregressive moving average* (ARMA), *autoregressive with exogenous input* (ARX) e outros métodos Box-Jenkins têm sido amplamente utilizados há muito tempo para a previsão de séries temporais, na área financeira, de controle e outras aplicações (MAKRIDAKIS *et al.*, 2008). Na previsão de geração eólica, vêm sendo aplicados com êxito (FUKUDA *et al.*, 2001; MILLIGAN *et al.*, 2004; LYDIA *et al.*, 2016; SIM *et al.*, 2019), especialmente para previsões de curto prazo (0 a 6 horas à frente).

O modelo ARMA é usado para a previsão de potência elétrica gerada por parques eólicos nos EUA (MILLIGAN *et al.*, 2004). Alguns dos modelos oferecem melhorias em relação ao modelo de persistência. O desempenho do ARMA é altamente dependente dos parâmetros do modelo.

O *improved time series method* (ITSM) baseado no ARIMA usa decomposição *wavelet* que é aplicada a uma subsérie de velocidade do vento para obter a previsão final da velocidade do vento. Os resultados da simulação indicam que o método proposto em Liu *et al.* (2010) melhora a precisão das previsões em relação ao modelo de séries temporais clássico, e ao de rede neural.

O modelo *autoregressive conditional heteroscedastic* (ARCH), é combinado com o modelo ARIMA considerando a heteroscedasticidade da série residual. O ARIMA-ARCH é utilizado para prever a subsérie de velocidade do vento obtido a partir da sua decomposição (*wavelet*). A previsão final da velocidade do vento é a soma desses valores de previsão. Os resultados mostram que ele pode melhorar a precisão da previsão (LV; YUE, 2011). O modelo ARIMA-ARCH é novamente utilizado para prever a velocidade do vento por Wang *et al.* (2012), e mostra-se que o ARIMA-ARCH apresenta melhor desempenho do que o ARIMA.

Para lidar com o teor não-estacionário de velocidade do vento, os dados de velocidade do vento são transformados para uma distribuição de Gauss, e normalizados para remover o teor não-estacionário por Zhang *et al.* (2011). O modelo paramétrico *autoregressive* (AR) é então aplicado aos dados de velocidade de vento, e é atualizado a cada hora durante a operação em tempo real, por um algoritmo recursivo dos mínimos quadrados. O método foi validado usando os dados de velocidade de vento de um parque eólico.

Em resumo, abordagens estatísticas convencionais são baseadas em modelos estatísticos lineares clássicos como o AR, *moving average* (MA), ARMA, e na abordagem de Box-Jenkins, onde Box-Jenkins é baseado em um modelo ARIMA ou um modelo ARIMA ajustado sazonalmente, do inglês *seasonal autoregressive integrated moving average* (SARIMA). No entanto, as referências mostram que a precisão da previsão pode ser melhorada, dependendo dos parâmetros selecionados para o modelo (JUNG; BROADWATER, 2014).

2.3.2.2 Abordagem estatística não convencional

Uma alternativa aos modelos clássicos estatísticos são os modelos baseados em técnicas de IC (COSTA *et al.*, 2008; LEI *et al.*, 2009; FADARE, 2010; MORALES *et al.*, 2010).

Nos modelos baseados em ANN, *multi-layer perceptrons* (MLP) (LIN *et al.*, 1996; ALEXIADIS *et al.*, 1998; KRETZSCHMAR *et al.*, 2004), *radial basis function* (RBF) (BEYER *et al.*, 1994) e redes neurais recorrentes (KARINIOTAKIS *et al.*, 1996; MORE; DEO, 2003), *small-world neural network* (WANG *et al.*, 2019), SVM (PAWAR *et al.*, 2019), lógica *fuzzy* (KHORRAMDEL *et al.*, 2018; LIU *et al.*, 2019), e ELM (LI *et al.*, 2019; ACIKGOZ *et al.*, 2020) têm sido usadas para esse fim.

A ANN é uma abordagem estatística amplamente utilizada para as previsões da velocidade do vento e de potência gerada. É constituída por uma camada de entrada, uma ou mais camadas escondidas, e uma camada de saída. Cada camada tem um número de neurônios artificiais, e utiliza uma abordagem conectiva para conectar os neurônios àqueles da camada anterior. Esta abordagem é capaz de modelar a relação não linear complexa entre as camadas de entrada e saída, através de um processo de aprendizagem. Esta abordagem não requer expressões matemáticas explícitas utilizadas nas abordagens físicas e estatísticas convencionais. Além disso, tem a capacidade de autoaprendizagem, auto-organização e auto-adaptação.

A ANN é uma tecnologia promissora na previsão da velocidade do vento. Lapedes e Farber (1987) propuseram um modelo de ANN do tipo *feed-forward* utilizando o algoritmo de aprendizado *backpropagation* (BP) desta rede na previsão da velocidade do vento. Song (2000) desenvolveu uma ANN que utiliza atrasadores e que a soma dos pesos das entradas deve ser igual a 1. O modelo apresentou bom resultado para previsões da hora seguinte quando os dados do vento não oscilam abruptamente. Outras técnicas são aplicadas para melhorar o resultado das ANN, Li *et al.* (2011) combinam o resultado de várias redes para realizar a previsão.

Algumas publicações comparam modelos de ANN com o modelo de persistência, e foi encontrado que normalmente as ANN se sobressaem em relação ao modelo de persistência, mas que algumas vezes essa vantagem não é muito significativa (SFETSOS, 2000; COSTA *et al.*, 2008; LEI *et al.*, 2009).

Jursa (2007) utilizou uma ANN e o PSO para a previsão de energia eólica de curto prazo. O PSO é utilizado para selecionar a variável de entrada a partir de vários locais próximos.

Amjady *et al.* (2011) propuseram um método de previsão composto por um PSO modificado (*enhanced particle swarm optimization* (EPSO)) e por uma rede neural híbrida modificada (*modified hybrid neural network* (MHNN)). O método composto apresenta alta capacidade de aprendizagem, e é capaz de evitar os problemas de *over-fitting* e mínimos locais. A técnica de seleção de recursos chamada de informação mútua (*mutual information* (MI))

é aplicada para selecionar as entradas candidatas mais informativas, filtrando entradas tanto irrelevantes, quanto redundantes para o método de previsão. Os resultados confirmam a validade da abordagem desenvolvida.

Blonbou (2011) faz uma abordagem com uma ANN em conjunto com um método de aprendizagem adaptativa, e uma aproximação Gaussiana é proposta para previsão de energia eólica em curtíssimo prazo. Este método é interessante por apresentar além da previsão pontual um intervalo de confiança desta previsão.

A *generalized feedforward neural network* (GFNN) é usada por Celik e Kolhe (2013) para prever a distribuição anual de densidade de probabilidade da velocidade do vento. Os mesmos parâmetros utilizados pela função de Weibull são utilizados como entradas. O método proposto apresenta melhor precisão para prever a distribuição de densidade de probabilidade da velocidade do vento do que a função Weibull.

Douak *et al.* (2013) apresentaram o algoritmo *kernel ridge regression* (KRR) com aprendizagem ativa, usado para a previsão da velocidade do vento. O aprendizado ativo atua como um seletor de amostra inteligente capaz de produzir modelos de previsão compactos, e também para filtrar amostras de treinamento ruidosas. Três métodos de aprendizagem ativa para a construção do conjunto de treinamento do KRR são propostos: associação de regressores (*pool of regressors active learning* (PAL)), distância do conjunto de treinamento mais próximo (*distance active learning* (DAL)), e regressão residual (*residual regression active learning* (RSAL)). Os resultados experimentais mostram que a coleta inteligente de amostras de treinamento pode ser benéfica para a previsão da velocidade do vento.

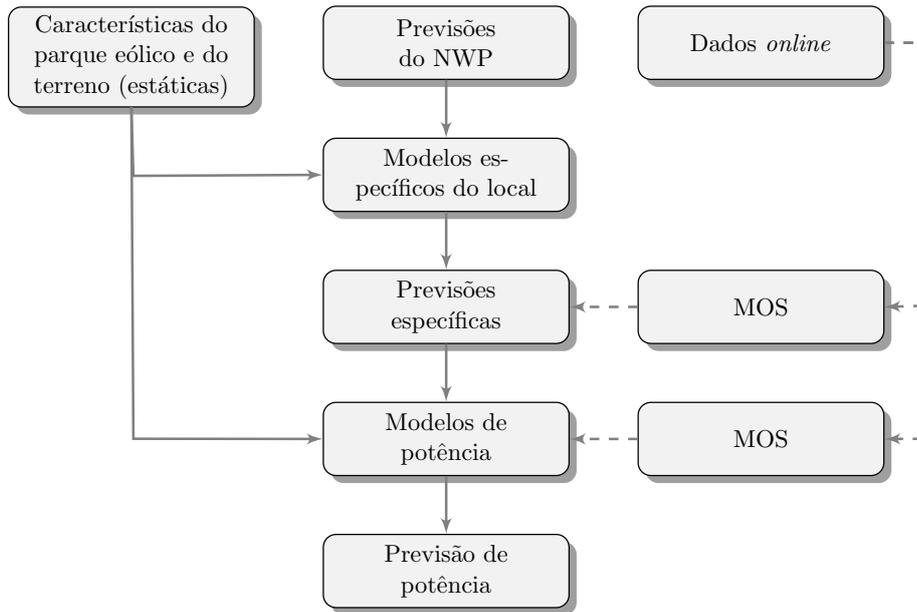
As arquiteturas das ANN utilizadas nesta tese são de aprendizado supervisionado, utilizando os métodos BP, DE, PSO ou o *particle swarm optimization with aging and weakening factor* (PSO-AWF). As ferramentas de IC e suas parametrizações serão apresentadas de forma mais detalhada nos Capítulos 3 e 4.

2.3.3 Modelos físicos

A abordagem física para a previsão, em contraste com a estatística, usa a descrição física detalhada para modelar as condições no local do parque eólico (KARINIOTAKIS *et al.*, 2004; LANGE; FOCKEN, 2006; GIEBEL; KARINIOTAKIS, 2017). A Figura 2 ilustra a operação básica de uma abordagem física.

As abordagens físicas são baseadas nos modelos usando os princípios físicos fun-

Figura 2 – Diagrama de blocos do sistema baseado na abordagem física.



Fonte: Adaptado de Kariniotakis *et al.* (2004).

damentais para a conservação de massa, momento, e energia em fluxos de ar. Embora existam muitos modelos CFD disponíveis, todos eles são baseados nos mesmos princípios físicos básicos. Eles diferem na forma como as grades são estruturadas, dimensionadas, e como os cálculos numéricos são executados.

Os modelos físicos permitem a previsão de parâmetros meteorológicos relevantes, que influenciam na previsão eólica, enquanto os modelos estatísticos são empregados para modelar a conversão dos parâmetros meteorológicos em potência, também aplicados na correção/refinamento da dinâmica de um local específico, entre outras aplicações.

Os métodos físicos utilizam características dos aspectos físicos do local onde ocorre a previsão, como topografia, rugosidade, obstáculos e a descrição das instalações do parque eólico (número de aerogeradores, *layout* dos aerogeradores, curva de potência, entre outros), para alcançar a melhor previsão possível e estimar a velocidade do vento no local do parque eólico. Uma maneira de lidar com a complexidade do terreno é o uso de mapas de alta resolução da área do parque no NWP (KARINIOTAKIS *et al.*, 2004; CUTLER, 2006).

Uma curva de potência é então usada para converter a previsão da velocidade do vento em previsão de potência gerada. A maioria dos modelos utiliza modelos estatísticos de saída, do inglês *model output statistics* (MOS), que é uma correção estatística para reduzir erros sistemáticos remanescentes. No caso do MOS ser utilizado, os métodos também dependem de medições da produção passada (SIEBERT, 2008). As entradas básicas de um modelo físico são

(KARINIOTAKIS *et al.*, 2004):

- Informação dinâmica: os dados do NWP para as horas seguintes fornecidos por um serviço meteorológico; algumas vezes são utilizadas medições *online* para a adaptação do MOS;
- Informação estática: descrição das instalações do parque eólico (*layout* e número das turbinas eólicas, curva de potência, etc.); descrições do terreno incluindo orografia (estudo das nuances do relevo de uma região), rugosidade, e obstáculos.

Um número de abordagens físicas estão disponíveis na literatura e no mercado (KARINIOTAKIS *et al.*, 2004; COSTA *et al.*, 2008; SIEBERT, 2008; LEI *et al.*, 2009; GIEBEL; KARINIOTAKIS, 2017). O Prediktor foi desenvolvido pelo laboratório nacional Risø na Dinamarca. Ele usa os programas *wind atlas analysis and application program* (WAsP) e o Park2 (RATHMANN *et al.*, 2018), para considerar as condições locais através das previsões de alta resolução de um NWP, o *high resolution limited area model* (HIRLAM) (LANDBERG, 1999a; LANDBERG, 1999b; LANDBERG, 2001). O Previento, desenvolvido na universidade de Oldenburg, na Alemanha, tem uma abordagem física semelhante, mas usa as previsões de um NWP diferente, o Lakelmodell (LocalPred), utilizado pelo serviço meteorológico alemão (FOCKEN *et al.*, 2001). O LocalPred foi desenvolvido no centro nacional de energias renováveis (CENER) na Espanha. Este modelo envolve a otimização adaptativa da previsão do NWP, modelagem de séries temporais, modelagem de meso-escala com o MM5, e modelagem da curva de potência (MARTÍ *et al.*, 2003).

O eWind, desenvolvido pela AWS TrueWind Inc. nos EUA, tem uma abordagem semelhante ao Prediktor, mas utiliza um modelo de camada limite de alta resolução (ForeWind) como um modelo numérico para levar em consideração as condições locais (MCCARTHY *et al.*, 2003).

Modelos físicos têm aplicações práticas limitadas no uso para previsões de curto ou curtíssimo prazo por problemas de latência e condições iniciais dos parques precisas que nem sempre podem ser garantidas (LIU; CHEN, 2019).

2.3.4 Modelos híbridos

Várias abordagens têm sido desenvolvidas com a combinação de diferentes modelos. A ideia básica dos híbridos é de combinar abordagens, extraíndo o melhor de cada modelo para obter uma performance global para o horizonte desejado. As combinações dos modelos podem ser baseadas em pesos, técnicas de pré-processamento de dados, seleção de parâmetros, técnicas

de otimização e de processamento de erros (TIAN *et al.*, 2018; SINGH *et al.*, 2019; MISHRA; DASH, 2019; QU *et al.*, 2019; TIAN *et al.*, 2020), abordagens baseadas em decomposição que tomam vantagem da decomposição de séries temporais, têm sido exploradas com mais frequência. A série temporal original é decomposta em diferentes subséries e modelada de maneira mais efetiva que a série original (LIU *et al.*, 2020). Segundo Kariniotakis *et al.* (2004), as combinações ainda podem ser classificadas como:

- Combinação de abordagens físicas e estatísticas (p.ex. Zephyr);
- Combinação de modelos de curto prazo (0 à 6 horas) e médio prazo (0 à 48 horas) (p.ex. MORE-CARE);
- Combinação de modelos estatísticos alternativos (p.ex. Sipleolico).

A combinação pode ser feita utilizando critérios de horizonte depois de identificado qual modelo é o melhor para aquele horizonte (p.ex. MORE-CARE), ou por um processo de seleção baseado na performance recente de cada modelo individual (p.ex. Sipleolico) (KARINIOTAKIS *et al.*, 2004).

O projeto ANEMOS utilizou um método chamado de *adaptive exponential combination* (AEC) (MARTI *et al.*, 2006). No primeiro passo, são utilizados vários métodos de combinação, sendo o AEC um deles. No segundo passo, o método AEC é usado para combinar as alternativas do primeiro passo.

Bouzgou e Benoudjit (2011) propuseram um sistema de arquitetura múltipla (*multiple architecture system* (MAS)) para a previsão da velocidade do vento. Diferentes algoritmos de regressão são considerados como modelos candidatos para fazer a previsão do conjunto: regressão linear múltipla (*multiple linear regression* (MLR)), ANN, RBF e SVM. Três estratégias de combinação são exploradas: média simples, média ponderada, e a fusão não-linear, por meio de um método de ANN. As estratégias de combinação propostas melhoram o desempenho em comparação a uma única previsão.

Zhang *et al.* (2012) propuseram dois tipos de modelos de previsão estatísticos univariados baseados nos mínimos quadrados (*least squares* (LS)). O modelo *least squares SVM* (LS-SVM) univariado e um modelo híbrido de ARIMA com o modelo LS-SVM. A abordagem híbrida é projetada para capturar tanto a componente linear, através do ARIMA, quanto os padrões não-lineares modelados pelo LS-SVM. No entanto, não houve melhorias significativas entre o LS-SVM e o modelo híbrido.

2.3.5 Modelos de previsão com correlação espacial

A previsão de correlação espacial é usada principalmente para a caracterização do recurso eólico no local onde a informação não é suficiente ou não está disponível, mas há uma estação de medição vizinha. É um indicador útil para avaliar o potencial eólico em áreas sem medições de vento.

Existem várias abordagens, mas a mais comum é o método *measure-correlate-predict* (MCP) (THØGERSEN *et al.*, 2007), que é amplamente utilizado para os estudos de integrações eólicas futuras. O método é utilizado para os dados de vento a longo prazo. Quatro métodos MCP diferentes são introduzidos: regressão linear, *matric*, Weibull e o modelo *wind index*. Verificou-se que todos têm bom desempenho.

Kwon (2010) usou o método MCP para prever a velocidade do vento a longo prazo, em seguida, uma simulação numérica (Monte Carlo) é executada para utilizar os modelos de probabilidade dos parques maiores. Este modelo é utilizado para avaliar a incerteza do potencial de energia eólica global antes da construção de turbinas eólicas.

O modelo hierárquico bayesiano é desenvolvido para caracterizar os dados de vento no local desejado usando recursos conhecidos (MIRANDA *et al.*, 2006). Possui duas etapas: os dados de velocidade do vento a partir do local de referência são definidos como a soma das componentes temporais, espaciais, e de erros na primeira etapa. Na segunda, um componente temporal e espacial é modelado como um *random walk* de primeira ordem e uma distribuição normal multivariada, respectivamente. Verifica-se que a inferência bayesiana pode ser uma ferramenta útil na previsão de correlação espacial.

Tastu *et al.* (2010) usaram um modelo autorregressivo de vetor paramétrico condicional para levar em consideração as outras medições regionais no oeste da Dinamarca e encontrou uma redução no RMSE para a previsão de uma hora à frente de até 18,5%. Tastu *et al.* (2011) analisaram correlações automáticas e cruzadas dos erros de previsão entre cinco regiões no oeste da Dinamarca e descobriram que “existe, em geral, uma correlação cruzada significativa entre os erros de previsão para áreas vizinhas com atrasos de algumas horas. Os atrasos com dependência significativa são de até 5 horas, enquanto os atrasos com maior efeito são os de 1 e 2 horas. Este padrão de correlação cruzada é claramente condicionado pela situação meteorológica predominante, caracterizada principalmente pela velocidade e direção do vento. A direção do vento desempenha um papel crucial”.

2.3.6 Modelos de previsão regionais (*Upscaling*)

Para prever a geração de energia eólica de uma região ou de um país é comum a soma das previsões da saída de cada parque eólico. Esta opção pode ser chamada de ‘força bruta’ porque prever cada uma das saídas de cada parque eólico em uma região/país pode ser muito caro e até mesmo proibitivo para o gerenciamento de dados e o esforço computacional (particularmente para a abordagem estatística). Além disso, os dados medidos pelo *supervisory control and data acquisition* (SCADA) (por exemplo, geração de energia eólica, velocidade do vento) de todos os parques eólicos, e o detalhamento destas plantas (potência nominal, curva das turbinas, entre outros), podem ser de baixa qualidade e às vezes até não existirem. Por exemplo, as informações *online* medidas pelos sistemas SCADA em alguns países não estão disponíveis, pois é obrigatório instalar o sistema apenas em grandes parques eólicos.

Para resolver este problema, as abordagens de *upscaling* foram desenvolvidas para prever a geração eólica regional/nacional a partir de uma amostra de parques eólicos de referência. Além disso, a agregação de parques parece reduzir o erro de previsão como resultado dos efeitos de suavização espacial (FOCKEN *et al.*, 2001; FOCKEN *et al.*, 2002; PINSON *et al.*, 2003). Este efeito de suavização é causado pela compensação das flutuações de produção dos parques eólicos por estar sendo prevista uma produção total de uma região.

2.3.7 Modelos de previsão probabilísticos

Os modelos de previsão discutidos até agora fornecem uma previsão pontual da velocidade do vento que pode ser esperada para um intervalo de tempo à frente. É importante saber não apenas a previsão pontual, mas também a sua incerteza. Conhecer a incerteza da previsão permite que o operador do sistema avalie os riscos da previsão pontual. Diferentes abordagens podem ser encontradas na literatura sobre a previsão probabilística.

2.3.7.1 Modelos de previsão probabilísticos paramétricos

A função Beta, do inglês *probability density function* (PDF), é usada para modelar uma PDF da previsão da geração eólica para o modelo de persistência (BLUDSZUWEIT *et al.*, 2008). Em primeiro lugar, os resultados previstos são divididos em classes de potência, ou caixas para modelar a distribuição de potência com a função Beta. O erro da previsão da função Beta é obtido somando os erros de cada distribuição.

A distribuição de Cauchy é proposta como um modelo de distribuição para o erro de previsão para o modelo de Persistência (HODGE; MILLIGAN, 2011). A forma da distribuição é modificada de forma significativa com a mudança do horizonte de previsão.

Há pouca literatura disponível sobre a abordagem paramétrica devido à dificuldade de definir a natureza estocástica da velocidade do vento.

2.3.7.2 Modelos de previsão probabilísticos não-paramétricos

Ao contrário da abordagem paramétrica, a abordagem não-paramétrica não se baseia em qualquer suposição sobre a distribuição.

Juban *et al.* (2007) propuseram uma abordagem não-paramétrica com base na estimativa da densidade do *kernel* para prever a distribuição da geração eólica. A distribuição da geração eólica é decomposta em uma parte contínua, correspondente a todas as ocorrências da velocidade do vento com produção nula, e uma parte discreta correspondente às probabilidades distintas de todos os valores de produção.

A estimativa da densidade do *kernel* é expandido para um modelo adaptativo temporal e adota diferentes *kernels* para diversos tipos de variáveis (BESSA *et al.*, 2012). O modelo de estimativa da densidade do *kernel* baseado no estimador Nadaraya-Watson é proposto para estimar a incerteza nas previsões da geração eólica de curto prazo. Ele adota *kernels* específicos para as variáveis explicativas e o desenvolvimento de um modelo adaptativo temporal. O modelo foi utilizado por Pinson *et al.* (2007) com dados reais de dois grandes parques eólicos nos EUA. Além disso, a estimativa da densidade do *kernel* baseado em outro estimador adaptativo temporal é apresentado por Bessa *et al.* (2012).

Pinson e Kariniotakis (2004) apresentaram uma metodologia genérica para a avaliação do risco da previsão. Em primeiro lugar, os intervalos de confiança com base na abordagem de reamostragem são calculados, considerando o horizonte de previsão, a classe de potência e o risco de *cut-off*. Além disso, o índice de risco meteorológico (*meteo-risk index* (MRI)), que avalia a estabilidade de tempo, é usado para ajustar os intervalos de confiança. O método proposto é genérico e pode ser aplicado a quaisquer modelos de previsão de vento. No entanto, não parece ter muita influência nas estimativas apresentadas por Nielsen *et al.* (2006a).

Previsões de energia eólica probabilísticos usando uma análise de regressão *local quantile regression* (LQR) é proposta por Bremnes (2004). Esta abordagem não requer que se aceite a distribuição como provável. Os resultados das simulações demonstram a sua eficácia.

Kou *et al.* (2012) formularam o modelo *warped gaussian process* (WGP) para fornecer previsões de probabilidade de curto prazo da geração de energia eólica. Ele converte uma série de energia eólica não gaussiana em uma série latente, a qual é bem modelada por um processo Gaussiano. Além disso, um método de dispersão é empregado para reduzir os custos computacionais do modelo.

Abordagens não-paramétricas são utilizadas quando não é possível formular a distribuição dos erros da previsão. Esta abordagem é interessante uma vez que é difícil definir a natureza estocástica do vento. Além disso, a velocidade do vento e a previsão de potência são muito diferentes, dependendo do horizonte de tempo de previsão e localização. Abordagens não-paramétricas podem ser aplicadas, neste caso, sendo adequadas para estimar as incertezas da velocidade do vento e geração.

2.3.8 Modelos de previsão marítimos

A geração eólica *offshore* se torna atrativa devido aos ventos mais frequentes e fortes. No entanto, a previsão de energia eólica *offshore* ainda está nas fases iniciais de desenvolvimento. Não existem muitos modelos de previsão específicos disponíveis para *offshore* na literatura. A maioria dos modelos de previsão foram concebidos para sistemas terrestres. Para a abordagem física da previsão no mar, as características meteorológicas da camada limite atmosférica marinha têm que ser consideradas, incluindo a estabilidade da camada, a interação das ondas com o vento, entre outros (KARINIOTAKIS *et al.*, 2003). No entanto, a abordagem estatística não precisa considerar as condições marítimas exatas. Os resultados dos vários métodos de previsão no mar são investigados por Tambke *et al.* (2006).

Tambke *et al.* (2005) realizaram a avaliação das previsões de vários NWP para um parque *offshore*. Para a previsão física, as características meteorológicas da camada limite atmosférica marinha são consideradas para calcular a velocidade do vento à altura do *hub* das turbinas eólicas.

Algumas abordagens estatísticas usam dados marítimos como uma entrada, mas as técnicas específicas para lidar com estes dados não são discutidas (BREMEN, 2006; SIDERATOS; HATZIARGYRIOU, 2007; PINSON *et al.*, 2009b; GALLEGO *et al.*, 2011; LIMA, 2015).

2.3.9 Ferramentas de previsão disponíveis

Nos últimos 40 anos, o desenvolvimento de ferramentas de previsão passa da fase pioneira à fase de pesquisa e, finalmente, à fase comercial. Alguns desses *software*, como o Prediktor, estão no mercado há mais de 30 anos. Alguns dos *software* novos são os resultados combinados dos anteriores, ou os usa como referência, e podem ser vistos na Tabela 2.

Tabela 2 – Lista de *software* de previsão de energia eólica

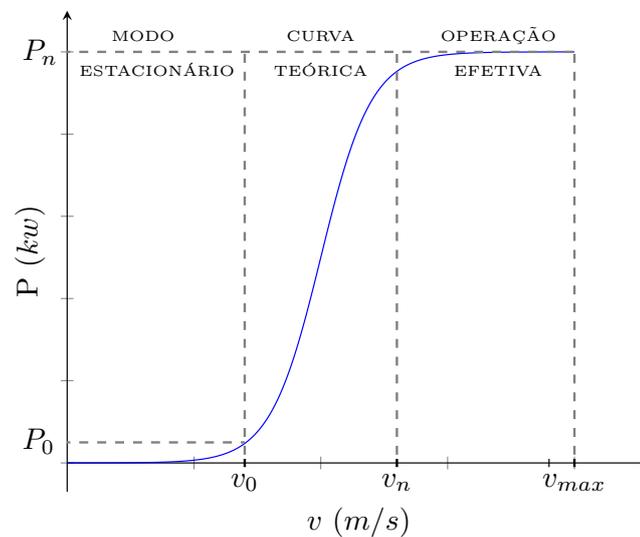
Ferramentas de previsão		
Modelo	Desenvolvedor	Método
AleaWind	AleaSoft, Spain	Estatístico
EPREV	INESC Porto/INEGI/CEsA/ CGUL, Portugal	Estatístico
GH Forecaster	Garrard Hassan, UK	Estatístico
PowerSight	3TIER, USA	Estatístico
RAL (More-Care)	RAL	Estatístico
SIPREÓLICO	University Carlos III, Madrid/ Red Eléctrica de España	Estatístico
Visionpoint	WindLogics Inc., USA	Estatístico
WPMS	ISET, Germany	Estatístico
WPPT	IMM; University of Copenhagen	Estatístico
Cassandra	University of Castilla-La Mancha/ Gamesa, Spain	Físico
HIRPOM	University College Cork, Ireland/ Danish Meteorological Institute	Físico
LocalPred-RegioPred	CENER	Físico
MeteoLógica	MeteoLógica, Spain	Físico
Precise Stream	Precision Wind, USA	Físico
Prediktor	Risø	Físico
Previento	University of Oldenburg, Germany	Físico
SOWIE	Eurowind GmbH, Germany	Físico
eWind	AWS Truewind Inc., USA	Híbrido
Scirocco	Aeolis Forecasting Services, Netherlands	Híbrido
WEFS	AMI Environmental Inc., USA	Híbrido
WEPROG	WEPROG, Germany	Híbrido
AWPT	ISET, Germany	Estatístico e ANN
AWPPS (More-Care)	Armines/Ecole des Mines de Paris	Estatístico e <i>Fuzzy</i> -ANN
Zephyr	Risø and IMM	Físico e Estatístico

Fonte: Adaptado de Soman *et al.* (2010) e Rankinen (2021).

2.4 Curva de potência de aerogeradores

Um aerogerador converte a energia cinética do vento em energia elétrica. Neste processo, a quantidade de potência que o aerogerador pode fornecer é diretamente dependente da velocidade e direção do vento entre 50 e 100 m do nível do solo. A Figura 3 apresenta a relação entre a potência de saída e a velocidade do vento que é modelada pela curva característica de potência de um aerogerador.

Figura 3 – Curva de potência típica de um aerogerador.



Fonte: Elaborado pelo autor (2021).

Por causa dos controles do processo de geração eólica, a conversão de energia eólica disponível no vento (que é proporcional ao cubo da velocidade do vento) em energia elétrica varia de forma não linear. A curva de potência pode ser dividida em quatro partes distintas. Para ventos entre 0 e 2-4 m/s a turbina não produz potência. A velocidade do vento que inicia a geração de energia elétrica é chamada de velocidade de partida (v_0). Entre a velocidade de partida e a de cruzeiro (v_n) que normalmente fica entre 12 e 16 m/s, a curva de potência apresenta uma inclinação ditada pela Equação 2.1 com potência gerada variando com o cubo da velocidade do vento (ACKERMANN *et al.*, 2005). A partir da velocidade nominal, a potência gerada é próxima à nominal em decorrência dos controles exercidos pelo aerogerador no sentido de limitar a potência extraída do vento. Em alguns casos, a potência máxima de saída de um aerogerador pode chegar a 20% a mais que sua potência nominal. Quando a velocidade chega à velocidade de corte (v_{max}), mecanismos de freio param o rotor da turbina, evitando assim estresses mecânicos

e possíveis danos à turbina.

$$P = \frac{1}{2} \rho_{ar} C_p A_r v^3, \quad (2.1)$$

em que, P é a potência da turbina eólica, ρ_{ar} é a densidade do ar, C_p é o coeficiente de performance da turbina, A_r é a área varrida pelo rotor da turbina eólica, e v é a velocidade do vento na altura do *hub*.

De acordo com a lei de Betz (1919), um aerogerador não consegue extrair mais que 16/27 (59,26%) da potência presente no vento. Na prática, essas taxas podem alcançar 45%, sem levar em conta perdas por conversões mecânicas. Os aerogeradores atuais extraem de 20% à 30% da potência no vento.

A previsão de potência gerada pode ser calculada direta ou indiretamente. No primeiro caso (direto), a previsão baseia-se em valores medidos de potência gerada e no segundo (indireto) faz-se necessário conhecer a curva de potência do aerogerador. A curva de potência pode ser obtida por testes realizados pelos fabricantes ou a partir de valores medidos em campo. Conhecida a curva de potência e a velocidade estimada do vento, a potência gerada poderá ser estimada, no caso indireto.

Modelar a curva de potência de um aerogerador não é tarefa simples. Efeitos como a turbulência e o sombreamento são comuns e devem ser considerados. O efeito de sombreamento² no parque pode ser evitado por um planejamento da localização das turbinas no parque.

2.5 Avaliação de performance dos modelos

Os algoritmos de previsão de geração eólica são caracterizados por uma incerteza inerente, o que significa que não existe previsão de geração disponível que seja exata. Portanto, é essencial que as previsões da geração eólica sejam avaliadas adequadamente, não só para avaliar o desempenho dos métodos escolhidos, mas também para obter uma compreensão mais profunda do que caracteriza a incerteza desta previsão.

Os critérios de verificação mais comuns são os níveis de erro ou faixas de erro comparando a previsão atual contra o valor observado na saída do parque eólico. O erro é a maneira mais comum de se avaliar uma previsão e existem vários métodos que podem ser

² O efeito de sombreamento ocorre quando a localização de uma turbina afeta no fluxo de vento de uma turbina que está atrás desta.

analisados (YOUSUF *et al.*, 2019). O erro da previsão é definido como

$$e_{t+\Delta} = P_{t+\Delta} - \hat{P}_{t+\Delta}, \quad (2.2)$$

em que, $e_{t+\Delta}$ é o erro da previsão, $P_{t+\Delta}$ é a potência medida, $\hat{P}_{t+\Delta}$ é a potência prevista. Geralmente o erro é normalizado, sendo dividido por P_n , que é a capacidade instalada do parque eólico. A normalização do erro permite a comparação dos erros de previsão relacionados aos parques de diferentes capacidades instaladas. Note que a Equação 2.2 pode ser generalizada, simplesmente considerando a variável de saída y .

$$e_{t+\Delta} = y_{t+\Delta} - \hat{y}_{t+\Delta}, \quad (2.3)$$

A verificação destes modelos não é nada trivial, pois depende da função custo envolvida. Vários indicadores de performance de modelos foram propostos e empregados na avaliação da previsão de velocidade do vento, mas nenhum é reconhecido como universal e padrão. Isso dificulta a comparação entre diferentes modelos, e como resultado é necessária a utilização de múltiplas medições.

As descrições de erro mais usuais são: tendência (BIAS), erro médio (*mean error* (ME)), erro médio absoluto (*mean absolute error* (MAE)), erro médio absoluto percentual (*mean absolute percentage error* (MAPE)), raiz do erro médio quadrático (*root mean square error* (RMSE)), histogramas da distribuição de frequência do erro, função de correlação, e valores do coeficiente de determinação (R^2). Na maior parte das vezes, o erro padrão é dado como um percentual da capacidade instalada, já que isto normalmente é o que mais interessa ao cliente (capacidade instalada é de fácil medição); algumas vezes o erro é dado pelo percentual da produção média ou em números absolutos (YOUSUF *et al.*, 2019).

O BIAS corresponde a uma estimativa do erro sistemático que é fornecido pelo erro médio sobre todo o período avaliado. Quando todo o conjunto é calculado, o BIAS fornece uma indicação de tendência positiva ou negativa da previsão, levando assim à identificação de uma tendência. Entretanto, é muito pouco provável que um método de previsão com uma tendência nula proporcione boas previsões, porque o BIAS cancela os resultados positivos com os negativos.

$$BIAS = \frac{1}{N} \sum_{t=1}^N e_{t+\Delta}, \quad (2.4)$$

O MAE mede a magnitude média do erro em um conjunto de previsões. Para ser mais específico, é a média sobre a amostra de verificação dos valores absolutos da diferença entre

o valor observado e a previsão correspondente. O MAE é uma medição linear, o que significa que todas as diferenças estão recebendo o mesmo peso na média.

$$MAE = \frac{1}{N} \sum_{t=1}^N |e_{t+\Delta}|, \quad (2.5)$$

O MAPE é a diferença entre o valor atual e o previsto dividido pelo valor atual. O módulo deste cálculo é então somado para cada previsão e dividido pelo número total de pontos previstos.

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{e_{t+\Delta}}{y_t} \right|, \quad (2.6)$$

O RMSE é uma medição quadrática, que mede a magnitude média do erro. Como os erros são elevados ao quadrado antes de ser calculada a média, o RMSE fornece um peso maior aos erros maiores. Isso significa que o RMSE é mais útil quando grandes erros são particularmente indesejáveis.

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (e_{t+\Delta})^2}, \quad (2.7)$$

Lange (2003) demonstrou como o RMSE pode ser decomposto em 3 valores separados: a tendência, a tendência do desvio padrão e a dispersão. O componente da dispersão descreve a contribuição das faixas de erro sobre o RMSE total. Outra possibilidade é avaliar a consistência dos cálculos da incerteza comparando com o observado. Para a verificação da previsão de eventos extremos, existem outros métodos como a probabilidade de detecção (*probability of detection* (PoD)) e taxas de alarme falso (*false alarm rate* (FAR)) (CUTLER, 2006).

2.5.1 Decomposição do RMSE da previsão da velocidade do vento

O RMSE pode ser decomposto em três partes diferentes, de acordo com o erro original da previsão (HOU *et al.*, 2001; MONTEIRO *et al.*, 2009)

$$RMSE^2 = BIAS^2 + SDE^2, \quad (2.8)$$

$$RMSE^2 = BIAS^2 + SDBIAS^2 + DISP^2, \quad (2.9)$$

$$SDE = \sigma(e), \quad (2.10)$$

$$SDBIAS = \sigma(v) - \sigma(\hat{v}), \quad (2.11)$$

$$DISP = \sqrt{2\sigma(\hat{v})\sigma(v)(1-r)}, \quad (2.12)$$

em que, SDE é o desvio padrão do erro, $\sigma(\hat{v})$ e $\sigma(v)$ são respectivamente o desvio padrão da previsão e da medição, e r é o coeficiente de correlação cruzada entre as duas séries temporais \hat{v} e v .

Devido à não-linearidade da função de transferência, os erros de previsão da velocidade do vento são penalizados de forma heterogênea, quando se pensa em termos de potência real. Erros em velocidades muito baixas são irrelevantes uma vez que a saída da geração é zero ou baixa. Os erros na região plana da curva de potência (entre 12 e 25 m/s) também são irrelevantes uma vez que a saída da geração é constante, a menos que se tenha de considerar mudanças na direção do vento ou na intensidade da turbulência. Erros entre velocidades baixas a moderadas (3-12 m/s) são altamente penalizados, pois um pequeno erro na velocidade leva a um grande erro na geração. No entanto, os piores erros são obtidos perto da faixa de corte do aerogerador (25 m/s), quando o sistema muda abruptamente de um valor máximo de saída para zero, ou vice-versa.

2.5.2 Avaliação da performance dos modelos probabilísticos

Avaliar as previsões pontuais clássicas basicamente consiste em avaliar o desvio ou discrepância entre a previsão e o valor real, conforme apresentado na Seção 2.5.1. Avaliar previsões de probabilidade é no entanto mais difícil. Na verdade, uma previsão probabilística não pode ser considerada como incorreta. O exemplo a seguir ilustra esta afirmação: a previsão probabilística indica que a geração de energia esperada para um determinado horizonte é entre 1 e 1,6 MW, com 50% de probabilidade, sendo o resultado real de 0,9 MW. A previsão probabilística cobre apenas 50% dos casos, e não é possível dizer se neste caso em particular, ela pertence ou não aos casos perdidos pelos intervalos de previsão.

Portanto, um método específico para avaliar previsões probabilísticas de geração eólica deve ser concebido. Pinson *et al.* (2006) apresentaram uma abordagem de avaliação que consiste em uma série de previsões de quantis para proporções nominais e observações (valores medidos) únicas ou variáveis. A classificação apresentada pode ser incondicional, mas como diversas variáveis podem influenciar a qualidade dos intervalos, a avaliação também pode se tornar condicional a fim de revelar a influência de tais variáveis.

Existem várias estruturas para a avaliação de previsões de intervalo na área econômica (WALLIS, 2003; CLEMENTS, 2005), que são baseadas em testar a hipótese de

cobertura condicional correta dos intervalos de previsão. Esse método foi introduzido para testar os intervalos de previsão de um passo à frente. Isto é equivalente a testar a cobertura incondicional correta dos intervalos e sua independência.

No entanto, na previsão da geração eólica deve-se considerar as previsões com vários passos à frente, onde há uma correlação entre os erros de previsão. Essa correlação é principalmente o resultado da inércia na incerteza da previsão meteorológica.

2.6 Incerteza das previsões de geração eólica

De maneira geral, os modelos descritos até agora disponibilizam na sua maioria, previsões pontuais para um número de horas à frente. Apesar da performance destes métodos poder ser avaliada usando os critérios apresentados na Seção 2.5.1, esses critérios apenas dão ao usuário final uma ideia da performance média destes modelos. Situações em que o modelo é usado como ferramenta de decisão, saber que o modelo é normalmente bom não é suficiente. Por isso, os usuários finais têm expressado a necessidade por ferramentas que não só entreguem o valor de produção esperado, mas também uma estimativa da acurácia desta previsão. Em outras palavras, o usuário quer saber o quão confiável é essa previsão.

O planejamento e a operação dos sistemas de potência precisam saber a incerteza da previsão de energia para que possam preparar precauções caso sejam necessárias. Fatores que podem contribuir na incerteza são:

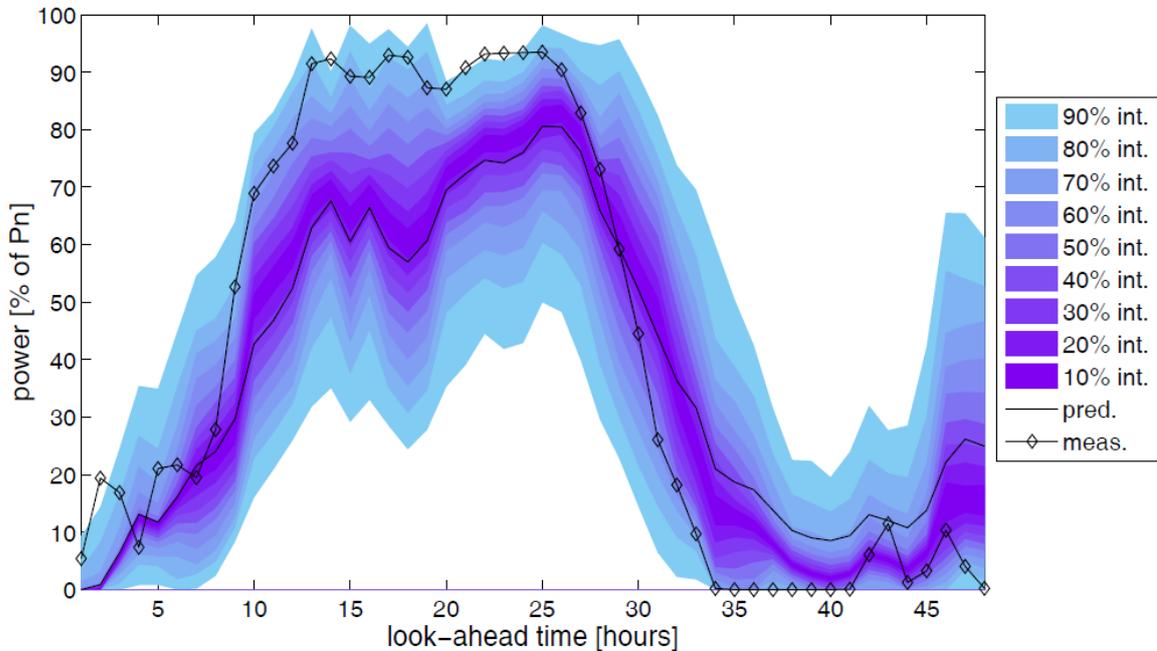
- Erros nas condições iniciais, medições de potência de saída, medições de parâmetros climáticos e nos dados climáticos de mesoescala obtidos pela simulação do modelo NWP;
- A previsibilidade do comportamento atmosférico é mais estável em algumas situações do que em outras. Como exemplo das menos previsíveis estão as tempestades, frentes e efeitos pontuais como insolações que causam o aquecimento do ar e a subida deste;
- Quanto mais complexo o terreno do parque eólico, maior será a turbulência no local, dificultando a previsão;
- A acurácia e o detalhamento utilizado no NWP, incluindo a resolução e o número de suposições feitas para simplificação;
- A precisão das curvas de potência das turbinas eólicas que definem a conversão entre a velocidade do vento e a potência. A curva de potência também muda com o tempo influenciada pela fadiga do material e incidentes naturais como o acúmulo de sujeira nas pás;

- A falha do gerador eólico, causado por inúmeros fatores de projeto e ou manutenção. A manutenção planejada das turbinas é de grande importância.

A incerteza pode ser trabalhada então de duas maneiras: uma como a faixa de confiança de uma previsão de energia eólica em um dado intervalo de tempo, e a outra é a variabilidade a qual a energia eólica poderia flutuar em um certo intervalo de tempo (CUTLER, 2006).

A primeira resposta para esta demanda proposta na literatura foi fornecer intervalos de previsão, os quais dão uma estimativa de quanto a produção poderá variar com um certo grau de confiança. Um exemplo deste tipo de previsão pode ser visto na Figura 4, onde os intervalos de previsão são representados por áreas coloridas.

Figura 4 – Previsões pontuais associadas a intervalos de confiança.



Fonte: Pinson (2006).

Um modelo paramétrico foi proposto por Lange (2005), que modela a distribuição condicional do erro da previsão da velocidade do vento (dada a velocidade do vento prevista) como uma distribuição gaussiana. A curva de potência é usada para transformar a distribuição gaussiana do erro da velocidade do vento em uma distribuição não gaussiana do erro da potência prevista. Os intervalos de previsão podem ser então derivados das distribuições dos erros da potência estimada.

Pinson (2006) propôs um algoritmo não-paramétrico que estima os intervalos oriundos de um conjunto de valores de erros passados. Os intervalos são computados usando distri-

buições condicionais do erro da previsão de potência (dada a velocidade do vento prevista). O erro e os dados da velocidade do vento os quais as distribuições são calculadas são determinados usando um sistema de inferência *fuzzy*. Uma vez que estes dados são selecionados, um método de reamostragem (*resampling*) é utilizado para determinar a distribuição desejada. A vantagem deste método sobre os métodos paramétricos propostos anteriormente é que esse não depende da hipótese do tipo de distribuição do erro.

Outra classe de métodos não-paramétricos foi proposta para intervalos de previsão: regressão quantil. Como o nome diz, o método estima quantis predefinidos de um conjunto de dados de entrada. Para mais detalhes sobre o assunto sugere-se a leitura de Nielsen *et al.* (2006a) e Bremnes (2004). De maneira resumida temos que a incerteza na previsão da geração eólica pode ser:

- previsões probabilísticas, que consistem em estimar a incerteza da previsão eólica que possa ser expressa como uma medida de probabilidade. Uma caracterização mais completa da distribuição condicional pode ser encontrada na literatura sobre previsões probabilísticas (PINSON, 2006);
- índices de risco, que fornecem informações sobre o nível de precisão esperado da previsão. Reflete a dispersão de um conjunto de previsões para um único intervalo de previsão em um determinado período previsto (PINSON *et al.*, 2009b); e
- cenários de geração, que contêm informações sobre as funções densidade de probabilidade para o horizonte de tempo, e também fornece informações sobre o desenvolvimento dos erros da previsão durante o intervalo, e a distribuição espacial (PINSON *et al.*, 2009a).

2.6.1 Índices de risco da previsão

A estimativa da incerteza na previsão da geração eólica é um assunto complexo que depende de vários fatores (condições meteorológicas, o efeito da suavização espacial dos parques eólicos, o nível de potência previsto). Uma situação atmosférica instável pode levar a um grande erro na previsão da velocidade do vento e, assim, as previsões de energia eólica serão ruins.

Por outro lado, as previsões mais precisas são esperadas durante períodos estáveis. Este comportamento não depende do método de previsão que é usado. Além da incerteza da previsão, por exemplo um conjunto de quantis, e a previsão pontual, é possível acessar os índices de risco que fornecem informações importantes sobre o nível esperado de precisão das previsões (ou seja, a previsibilidade da situação atmosférica). Dois índices de risco foram propostos na

literatura

- MRI que reflete a dispersão do conjunto NWP disponível em um determinado momento (PINSON; KARINIOTAKIS, 2004); e o
- índice normalizado de risco da previsão (*normalized prediction risk index* (NPRI)), que reflete a dispersão de um conjunto de previsões da geração eólica para uma única previsão, ou para um período previsto (PINSON *et al.*, 2009b).

O índice de risco da previsão consiste de um único valor numérico (ou valor qualitativo), que fornece um aviso a priori sobre o nível esperado de erro da previsão. O índice de risco pode ser mais facilmente compreendido pelos operadores e utilizadores de previsão, em comparação com as previsões de probabilidade, e, ao mesmo tempo, o índice pode também estar relacionado com a informação sobre o valor da magnitude potencial dos erros da previsão. A partir desses índices de risco, é possível entender o quão preciso é o erro da previsão do vento. Os índices de risco podem ser utilizados para ampliar ou restringir intervalos previstos. Por exemplo, se o MRI é baixo, é esperado que o modelo seja preciso.

O MRI utiliza conjuntos (*ensembles*) de NWP de dois tipos: (i) aqueles obtidos pela perturbação das condições iniciais do modelo NWP, ou por diferentes modelos NWP; e (ii) previsões defasadas obtidas a partir de diferentes condições iniciais, no entanto, com um modelo não perturbado. Estas previsões são para os mesmos intervalos, mas elas são feitas em diferentes instantes de tempo. A dispersão é medida por duas-normas para calcular a distância entre as duas previsões.

Os NPRI são derivados das previsões de um conjunto NWP convertido em geração eólica por um modelo que relaciona vento e potência, tal como o descrito e avaliado por Nielsen *et al.* (2006b). Esse índice mede a dispersão do conjunto de geração eólica em um determinado período de tempo, ou para cada janela de previsão. A dispersão é calculada com um desvio padrão ponderado dos membros do conjunto. Os pesos são interpretados como a capacidade de cada membro do conjunto fornecer informações sobre a previsibilidade.

Pinson *et al.* (2009b) estabeleceram uma relação entre várias classes igualmente povoadas de valores do NPRI (5 classes), e distribuições de desequilíbrio energético. Os autores mostraram que para diferentes tipos de conjuntos de geração eólica, o NPRI pode fornecer informações úteis sobre o nível esperado de incerteza da previsão. Estes índices de risco foram implementados no sistema de previsão *armines wind power prediction system* (AWPPS) e estimado *online*.

2.7 Valor da previsão da geração eólica

Apesar da importância da previsão da geração eólica ser indiscutível para facilitar a integração de grandes montantes de potência eólica no sistema de potência, os esforços empreendidos não oferecem respostas satisfatórias tornando-se necessários continuidade na investigação. Um dos motivos para isto acontecer é a singularidade do sistema de potência. Cada sistema tem um *mix* de geração diferente, em que a variação e a natureza incerta da geração eólica podem ser menos impactantes para sistemas que possuam muitas unidades de despacho de resposta rápida (caso brasileiro) do que sistemas que não os utilize entre outras características.

Em Milligan *et al.* (1995) foi feito um modelo do sistema de potência para verificar o impacto da acurácia da previsão da geração eólica nos custos operacionais do sistema. Os resultados deste estudo mostram que o aumento da acurácia leva a custos menores. Entretanto, a penalidade associada à sobre ou subestimar a previsão da geração eólica não é simétrica. O custo associado a um erro positivo não é igual ao custo associado para o erro negativo de mesmo módulo. Os autores encontraram que essa dissimetria pode variar de um sistema para outro. Para um sistema sobre-estimado foi menor o custo, enquanto para outro foi o contrário.

Devido à dificuldade de estimar diretamente a redução do custo operacional do sistema de potência com o erro da previsão da geração eólica, muitos estudos têm tomado uma abordagem um pouco diferente, estudando o valor da previsão como ferramenta para o mercado de energia.

Sorensen e Meibom (1999) conduziram um estudo no mercado da Nordpool, operador de sistema dos países nórdicos. Os autores simularam uma quantidade considerável de energia eólica produzida pela Dinamarca e examinaram seus impactos em Nordpool. Em outro estudo (PINSON, 2006; PINSON *et al.*, 2004; PINSON *et al.*, 2007) o caso do mercado de energia alemão APX foi examinado usando o modelo AWPPS desenvolvido por Kariniotakis *et al.* (1996) para fornecer previsão de geração eólica. Os autores encontraram que por usar o modelo sozinho, os ganhos representariam 87% dos ganhos que poderiam obter usando previsões perfeitamente precisas (previsões com erro zero).

Pinson e Kariniotakis (2004) propuseram o uso do método de avaliação da incerteza para desenvolver estratégias específicas de leilão que levem em conta a natureza assimétrica do desequilíbrio de preços. Usando este método foi obtido ganhos de até 96% dos ganhos possíveis (utilizando previsão com erro zero).

Os benefícios econômicos da previsão do vento podem ser vistos em estudo feito

pela GE Energy para a *New York state energy research and development authority* (NYSERDA) e o *New York independent system operator* (NYISO). O estudo foi uma avaliação dos efeitos da integração da energia eólica no planejamento do sistema de transmissão, segurança, e operação. Foi examinado o impacto econômico geral da integração de hipotéticos 3.300 MW oriundos de energia eólica no sistema do estado de Nova York. Os 3.300 MW projetados representam aproximadamente 10% do sistema durante o pico de carga. O estudo da GE calculou os preços de energia baseados em um modelo que funciona para previsões do dia seguinte, com uma hora de antecedência em relação aos mercados de energia. Isso é diferente da maioria das análises anteriores sobre o impacto econômico da produção de energia eólica, que consideraram apenas os custos operacionais. No estudo da GE assumiu-se que a energia eólica produzida é *price-taker*. Isso significa que o balanço de oferta e demanda no mercado determina o preço e que as unidades geradoras eólicas são pagas a preço de mercado (RICHARD, 2005).

A conclusão geral do estudo foi que o sistema de potência do estado de Nova York pode confiantemente acomodar pelo menos 10% de penetração da energia eólica em relação à carga de pico do sistema com apenas pequenos ajustes no planejamento, operação e práticas de confiabilidade. Além disso, este estudo estimou que os custos variáveis operacionais anuais (combustível, custos de *start-up* da planta, etc.) do sistema seriam reduzidos em 350 milhões de dólares com a adição da eólica. Também foi estimado que 125 milhões de dólares, ou 36%, da redução dos custos está associado com o estado-da-arte da previsão de potência eólica. Isso é cerca de 80% da redução de custo estimado que pode ser conseguido com uma perfeita previsão de produção de energia eólica (RICHARD, 2005).

2.8 Conclusões

O sistema de energia hoje enfrenta desafios crescentes para manter um fornecimento de energia seguro e confiável. Parte deste desafio é a possibilidade de níveis significativos de geração eólica (incerteza) serem instalados no sistema elétrico. Isso traz novos desafios que envolvem a gestão da intermitência do vento. A previsão da geração eólica é necessária para enfrentar esses desafios.

A previsão de energia eólica pode ser feita em vários níveis desde para o planejamento da manutenção de uma única turbina, até a previsão de geração de todo um país. Existem vários modelos no mercado, cada um com um foco, não existindo assim uma solução melhor ou pior, ou alguma ferramenta que atenda todas as necessidades em qualquer cenário.

A previsão de energia é uma ferramenta essencial para facilitar a integração de grandes blocos de geração eólica à rede elétrica. Também complementa outras tecnologias de integração como: compensação de linha e eletrônica de potência, baterias e outras tecnologias de armazenamento de energia, e integração com geração hidrelétrica.

Avanços recentes na informática, nos modelos NWP, na modelagem de mesoescala do fluxo de vento sobre terrenos complexos, e a aplicação de IC e outras técnicas estatísticas implicaram em um grande aumento da acurácia das previsões de curtíssimo, curto, médio e longo prazo.

De particular relevância desta revisão, foi que para horizontes de até 6 horas aproximadamente, métodos estatísticos usando informações locais são superiores aos modelos físicos, os quais requerem horas de esforço computacional, introduzem imprecisões resultantes dos dados oriundos dos NWP, e necessitam de um alto nível de conhecimento da região onde será aplicado.

3 DESCRIÇÃO DO PROBLEMA E FERRAMENTAS DE INTELIGÊNCIA COMPUTACIONAL

Espaços de características (*feature space*) de alta dimensão têm chamado atenção para a estimativa de funções não lineares. Uma aplicação direta dessa abordagem é a regressão, onde algum mapeamento não linear é seguido por processamento linear em um espaço de variáveis de alta dimensão.

Seja um parque eólico, em que (x_i, y_i) são pares de dados históricos, que representam as variáveis de entrada e saída do modelo, onde x_i pode ser medição de velocidade do vento, potência gerada, entre outros no instante t , e y_i é a medição da variável de interesse (velocidade do vento, potência gerada,...) no instante de tempo $t + \text{horizonte}$ (Δ). O problema consiste em estimar uma função potencialmente não linear $f(\cdot)$ que descreve a relação entre as variáveis de entrada e saída, denotada como $\hat{f}(\cdot)$. Essa função calcula a predição do parque eólico que pode ser uma única variável (\hat{y}), ou uma matriz com n predições (\hat{Y}) de um ou vários parques. Essas n predições podem ser de diferentes variáveis (velocidade do vento, potência gerada,...), como também pode ser a predição de uma única variável de diferentes parques. A função desconhecida $f(\cdot)$ pode mudar com o tempo, ao contrário das regressões padrões que assumem modelos estáticos. O objetivo é encontrar uma estimativa $\hat{f}(\cdot)$ de $f(\cdot)$ que minimize o indicador de erro utilizado, neste caso o RMSE (Equação 2.7).

A aproximação linear deste problema é dado por $\hat{f}(x_t) = Ax_t$, em que $A \in \mathbb{R}^{n \times m}$ (ou $\mathbb{C}^{n \times m}$) é uma matriz de coeficientes que serão determinados. Muitos métodos de estimação baseados nessa aproximação já foram estudados (BOLOIX-TORTOSA *et al.*, 2017; JIA *et al.*, 2019; WEI *et al.*, 2019).

Alternativamente, pode-se declarar a aproximação em termos de mapeamento de uma matriz de pesos sinápticos de uma ANN, que são calculados através de um algoritmo de aprendizado. Um exemplo destes algoritmos de aprendizado é o BP. A redução de erros através do gradiente descendente relacionado as ANN tem sido discutida desde a década de sessenta (KELLEY, 1960; BRYSON; DENHAM, 1962; AMARI, 1967; WILKINSON, 1971), que logo em seguida foi usado explicitamente para minimizar funções de custo adaptando parâmetros de controle (pesos) (DREYFUS, 1973). Um artigo que contribuiu significativamente para a popularidade do BP em ANN foi a publicação de Williams e Hinton (1986), que descreve muitos casos em que o BP gerava resultados em um menor intervalo de tempo que as outras abordagens da época, tornando assim as ANN viáveis na resolução de problemas antes inviáveis. Hoje, o BP

é considerado uma referência.

Embora o BP seja uma técnica de aprendizagem consolidada (Li-Chiu Chang *et al.*, 2012; MAO *et al.*, 2016), verificou-se que a aplicação do PSO como algoritmo de aprendizagem são muito eficientes (ARCE *et al.*, 2016; RAZA *et al.*, 2016). Por isso, o interesse em investigar a aplicação dessa ferramenta em ambientes dinâmicos.

Para a segunda abordagem desta tese, a aproximação é definida em termos de um mapeamento $\phi(\cdot)$ em um ambiente não-linear, escrevendo $\hat{f}(x_t) = A\phi(x_t)$ sendo $A \in \mathbb{R}^{n \times d}$. As propriedades dos *kernels* de Mercer (VAPNIK, 2013) tornam possível derivar esquemas de estimação para $f(\cdot)$ em um espaço d-dimensional sem realizar os cálculos em tal espaço. Isso combina uma implementação simples de métodos lineares com as propriedades vantajosas de trabalhar com o mapeamento não linear.

Para a organização dos princípios básicos destas ferramentas, foi feita uma separação em dois grandes agrupamentos: métodos *offline* onde os parâmetros dos modelos não mudam após a fase de treinamento e métodos *online* que os parâmetros mudam sempre que lhes são apresentados novas informações.

Este capítulo segue a seguinte organização: a Seção 3.1 apresenta as ANN, tipos de treinamento, e trabalhos relacionados; a Seção 3.2 introduz o algoritmo proposto, PSO-AWF, para métodos *offline*; a Seção 3.3 descreve o funcionamento básico das máquinas *kernel*, modelos utilizados como referência, e trabalhos relacionados; por fim, a Seção 3.4 encerra o capítulo apresentando o algoritmo proposto, *multiple challengers kernel recursive least-squares* (MC-KRLS), para métodos *online*.

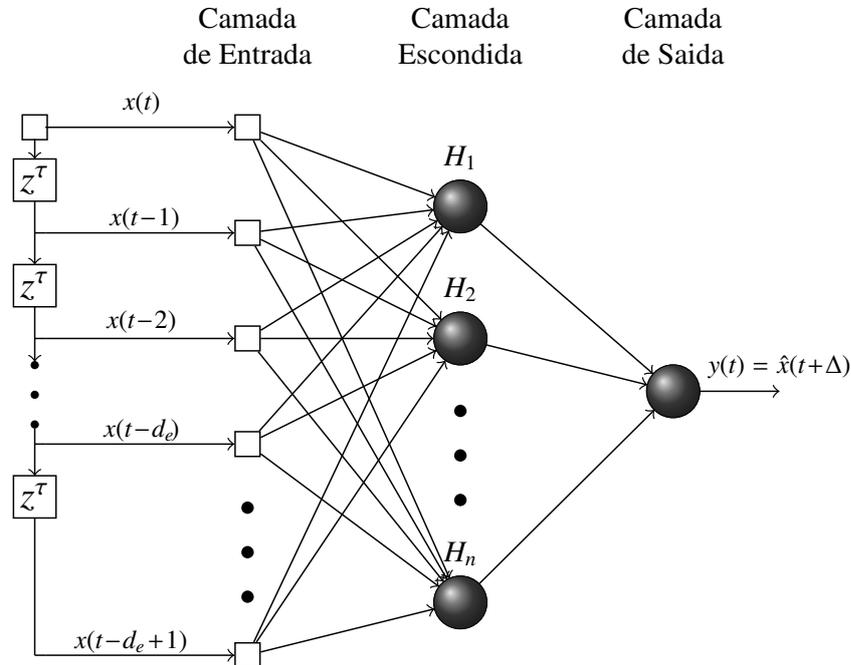
3.1 Ferramentas para o método *offline*

Neste trabalho apresentamos um algoritmo para a previsão de curtíssimo e curto prazo baseado em uma rede MLP com unidades de atraso, chamada FTDNN (PRINCIPE *et al.*, 1999). A FTDNN é uma rede multicamadas sem realimentação, seus pesos sinápticos podem ser ajustados com o BP. Como uma típica ANN *feed-forward*, o sinal passa pela FTDNN em apenas um sentido e direção, da entrada para a saída, e os neurônios de uma mesma camada não são conectados entre si. *Time Delay Neural Networks* (TDNN) são redes temporais que, junto com a entrada atual $x(t)$, têm uma linha de atrasos (LAG) inseridos na camada de entrada que representam a memória da rede, $x(t-1)$, $x(t-2)$, ..., $x(t-LAG)$ (ENGELBRECHT, 2007).

A Figura 5 apresenta uma FTDNN com d_e elementos da série contados a partir do

elemento atual $x(t)$, espaçados um do outro de τ unidades de tempo (neste exemplo com valor igual a -1), $y(t)$ é a saída do neurônio e $\hat{x}(t + \Delta)$ é a variável escalar prevista.

Figura 5 – FTDNN com arquitetura genérica.



Fonte: Elaborado pelo autor (2021).

O DE proposto por Storn e Price (1997) é utilizado como uma ferramenta de comparação além do BP e o PSO. O DE é uma estratégia de busca estocástica baseada em população para resolver uma ampla variedade de problemas de otimização. O desempenho de otimização do DE não depende apenas da escolha de quatro parâmetros de controle (ou seja, parâmetro de controle de mutação, *scaling factor* (F), parâmetro de controle de cruzamento, *crossover rate* (CR), e tamanho da população (NP)), mas também da escolha de estratégias de geração de vetores de teste (ou seja, mutação e estratégias de *crossover*). Um levantamento do estado da arte pode ser encontrado em Wu *et al.* (2021) e Saini e Saha (2021).

Apesar do BP ser uma técnica de aprendizado consolidada (BERGMEIR *et al.*, 2012; Li-Chiu Chang *et al.*, 2012; YAN, 2012; MAO *et al.*, 2016), é observada que a aplicação do PSO como um algoritmo de aprendizado é muito eficiente (BERGH; ENGELBRECHT, 2000; GUDISE; VENAYAGAMOORTHY, 2003; LIANG *et al.*, 2006; FEI *et al.*, 2009; ZHAO; YANG, 2009; WANG *et al.*, 2010; ARCE *et al.*, 2016; RAZA *et al.*, 2016). Por esta razão, surgiu o interesse em focar a investigação nesta ferramenta.

3.1.1 Particle swarm optimization

O PSO é uma técnica de otimização inspirada pelos comportamentos sociais cooperativos e competitivos de animais que se organizam em bandos e cardumes. O PSO pode ser utilizado para resolver problemas não lineares e multi modais, sendo assim apto a realizar diferentes buscas simultaneamente. O PSO mantém uma nuvem de partículas, onde cada uma destas representam uma possível solução. As partículas fluem através de um espaço multi-dimensional, onde a posição de cada partícula é ajustada de acordo com sua própria experiência e a de seus vizinhos.

Em um espaço de busca d -dimensional com uma solução ótima global, cada partícula é randomicamente posicionada, e suas respectivas posições $x_i(t)$ são inicialmente definidas como seus $Pbest$, que podem ser descritos como as suas melhores posições individuais até o momento. A melhor solução entre todas as partículas é o $Gbest$. Usualmente a velocidade inicial v de cada partícula é zero ou um valor randômico. A nova velocidade $v_i(t + 1)$ que orienta todo o processo de otimização, e é usada como uma ferramenta de comunicação entre todas as partículas, sendo descrita da seguinte maneira

$$v_i(t + 1) = v_i(t) + C_1 R_1 (Pbest_i(t) - x_i(t)) + C_2 R_2 (Gbest(t) - x_i(t)), \quad (3.1)$$

em que C_1 e C_2 são respectivamente as constantes sociais e cognitivas, R_1 e R_2 são valores randômicos entre $[0,1]$. A nova posição $x_i(t + 1)$ é então computada por

$$x_i(t + 1) = x_i(t) + v_i(t + 1). \quad (3.2)$$

Todas as partículas da nuvem tendem a se deslocar para o ponto ótimo. Durante a busca, se o $Gbest$ não é a solução ótima, uma das partículas pode encontrar uma solução melhor e se tornar o novo $Gbest$. Esse ciclo continua até que a solução ótima global ou uma restrição de parada do algoritmo é encontrada.

3.1.2 Trabalhos relacionados

Inúmeras variações do PSO tentaram eliminar algumas características do algoritmo clássico, como as partículas inativas, convergência rápida, e o risco do agrupamento em um mínimo local (ZHANG *et al.*, 2015). As variantes do PSO podem ser classificadas nas quatro categorias que seguem.

A primeira categoria de variações do PSO alcançaram melhorias através do ajuste de parâmetros. Shi e Eberhart (1998) foram os primeiros a introduzir um novo parâmetro, denominado por peso de inércia (ω) para influenciar a convergência. A atualização da velocidade mudou então para

$$v_i(t+1) = \omega v_i(t) + C_1 R_1 (Pbest_i(t) - x_i(t)) + C_2 R_2 (Gbest(t) - x_i(t)). \quad (3.3)$$

A redução linear de ω foi proposta por Shi e Eberhart (1999) e um método *fuzzy* para ajustar ω não linearmente também foi desenvolvido por Shi e Eberhart (2001). Além disto, os ajustes para outros parâmetros foram testados, mesmo assim as partículas continuam atraídas pelo *Gbest*. Caso o *Gbest* esteja preso em ótimo local, o problema de convergência prematura permanecerá.

A segunda categoria de algoritmos tem como objetivo o aumento da diversidade através da definição de topologias da vizinhança. Kennedy e Mendes (2002) introduziram a versão local do PSO. A regra de atualização da velocidade na Equação 3.3 substitui o *Gbest* pelo *Lbest*, que é a melhor posição encontrada por uma partícula em sua vizinhança, predefinida por topologia. Para tornar o PSO mais robusto, estruturas que variam no tempo também foram propostas por Suganthan (1999), Hu e Eberhart (2002), Beheshti *et al.* (2014) e Lim e Isa (2014). Chen *et al.* (2013) propuseram um controle adaptativo da vida útil do líder da nuvem. Caracterizado por uma idade crescente e uma vida útil, permitindo assim que outros indivíduos desafiem o líder quando ele se tornar velho. Em geral, ao ser introduzida vizinhanças, as variantes do PSO melhoram sua habilidade de prevenir uma convergência prematura. Contudo, como a atração ao *Gbest* é enfraquecida e outras partículas também podem influenciar os ajustes de velocidade e posição, esse tipo de variação do PSO tende a ser significativamente mais lenta.

A terceira categoria é a dos PSO híbridos com técnicas de buscas auxiliares para melhorar a performance. Vários tipos de paradigmas evolucionários foram usados para isso, incluindo o algoritmo genético (JUANG, 2004; GARG, 2016), evolução diferencial (ZHANG *et al.*, 2003; MAO *et al.*, 2017), otimização por colônia de formigas (HOLDEN; FREITAS, 2005; ELLOUMI *et al.*, 2014), entre outros. De maneira geral, integrar o PSO com outras técnicas de otimização é eficiente na melhora de sua performance, mas a desvantagem é que o algoritmo híbrido é usualmente bem mais complexo que o PSO original.

A quarta e última categoria introduz técnicas de múltiplos enxames com o intuito de aumentar a performance. O PSO original funciona melhor em espaços de busca de baixa dimensão, Bergh e Engelbrecht (2004) decompueram o espaço de busca e utilizaram múltiplos enxames para otimizar diferentes componentes de um vetor solução de maneira cooperativa.

No PSO darwiniano proposto por Tillett *et al.* (2005), um grupo de enxames de partículas paralelos colaboram para simular a seleção natural e escapar de ótimos locais. Liang e Suganthan (2005) propuseram o *dynamic multiswarm PSO* (DMS-PSO) que é caracterizado por pequenos enxames e mudanças na topologia de cada enxame. O *parallel comprehensive learning PSO* (PCLPSO) proposto por Gülcü e Kodaz (2015), múltiplos enxames baseados no paradigma mestre-escravo trabalham cooperativamente ou simultaneamente. Liu *et al.* (2017) propuseram uma nova técnica coevolucionária chamada de *cooperative multi-swarm PSO for dynamic multi-objective optimizations* (CMPSODMO). O número de enxames é determinado pelo número de funções objetivo, ou seja, cada função corresponde a um enxame. Usualmente, em um PSO com subpopulações múltiplas, as sub-populações podem ser tratadas como um tipo especial de vizinhança (LIANG; SUGANTHAN, 2005). Neste sentido, esta categoria de variantes do PSO também encaram problemas de lentidão na convergência.

Partindo das discussões anteriores, pode ser percebido que ainda é um desafio neste campo de pesquisa prevenir a convergência prematura. Enquanto retém a velocidade de convergência do PSO original, a maioria das variações alcançam a preservação da diversidade populacional ao custo da lentidão em sua convergência ou algoritmos complicados. O algoritmo proposto nesta tese, PSO-AWF, evita convergência prematura, reduzindo a velocidade de convergência e mantendo a simplicidade da estrutura do PSO. O PSO-AWF tem a habilidade de ‘saltar’ ótimos locais. Diferentemente das variações locais do PSO, o PSO-AWF preserva a informação do líder. Apenas quando o líder falha, as partículas divergem dele. Desta maneira, o PSO-AWF pode manter a rápida convergência característica do PSO.

3.2 Particle swarm optimization with aging and weakening factors (PSO-AWF)

Em colônias de animais, o envelhecimento facilita a substituição do líder por um indivíduo mais jovem, portanto cria assim mais oportunidades para diversificação e melhorias do grupo. Inspirado nesse fenômeno, este trabalho propõe uma variação do PSO onde outras características bio inspiradas são adotadas, o processo de envelhecimento e enfraquecimento.

O fenômeno envelhecimento tem atraído a atenção dos biólogos por aproximadamente 150 anos. Na década de 1880, Weismann (1882) primeiro argumentou que o envelhecimento é necessário para a evolução, já que provê um mecanismo para criar espaço para o desenvolvimento da próxima geração. Goldsmith (2006) defendeu que o envelhecimento proporciona o aumento da diversidade genética, e assim desempenha um papel importante na evolução

das espécies. A otimização da vida útil é de fundamental importância para a eficácia da evolução. Com os recentes desenvolvimentos na teoria do envelhecimento (GAVRILOV; GAVRILOVA, 2002; GOLDSMITH, 2004; GOLDSMITH, 2006), se torna interessante o estudo do impacto do envelhecimento sobre técnicas de inteligência computacional.

No PSO convencional a adequação dos indivíduos é determinada apenas pelo valor da sua função objetivo. Em sistemas naturais, a idade de um indivíduo também é importante ao determinar a sua adequação. A essência do PSO-AWF é prover um mecanismo que promova um líder adequado ao enxame através do envelhecimento. Como todas as partículas são atraídas pelo *Gbest* no PSO original, o *Gbest* pode ser visto como o líder da nuvem. O efeito enganoso gerado por um líder posicionado em um ótimo local é a causa principal da convergência prematura do PSO original. Para enfraquecer tal efeito, o PSO-AWF atribui um valor de vida útil para toda a nuvem. Similar ao PSO original, partículas ainda aprendem a partir do seu líder ao atualizarem suas velocidades e posições. Entretanto, se a vida útil da partícula chegou ao fim, novas partículas emergem para desafiar e tomar a liderança da nuvem. Dessa maneira, o mecanismo de envelhecimento provê oportunidades para outras partículas liderarem o enxame e assim trazer diversidade (BEZERRA *et al.*, 2017).

O envelhecimento não é a única característica que compõe a LE de um indivíduo. Neste trabalho são representadas todas as outras características da LE como o WF. O WF é utilizado toda vez que uma partícula (x_i) não obtém um novo resultado ($Pbest_i(t)$) melhor que o seu anterior ($Pbest_i(t - 1)$). Dessa maneira, mesmo partículas com a mesma idade terão diferentes LE e deixarão de existir em diferentes momentos. O algoritmo proposto é descrito nos passos que seguem

- *Inicialização*: LE, AF, WF e o número máximo de iterações são números inteiros definidos pelo usuário. As posições iniciais ($x_i(t)$) de todas as partículas são geradas randomicamente dentro do espaço de busca d-dimensional, com velocidades ($v_i(t)$) inicializadas em zero. A melhor partícula do enxame é selecionada como líder (*Gbest*). As idades das partículas (*PA*) são inicialmente nulas ($PA = 0$);
- *Atualização da velocidade e posição*: Todas as partículas seguem a regra de atualização da velocidade da Equação 3.1 e posição da Equação 3.2 para ajustar suas velocidades e posições;
- *Atualização do *Pbest* e *Gbest**: Para todas as partículas, se a posição recém gerada $x_i(t)$ é melhor que o seu $Pbest_i$, então $x_i(t)$ se torna seu novo $Pbest_i$. Adicionalmente, se a

melhor posição encontrada nesta iteração é melhor que o $Gbest$, então o $Gbest$ é atualizado para esta nova posição. Desta maneira, esse passo é semelhante ao do PSO convencional, mas neste caso sendo representado pelas melhores soluções encontradas pelas partículas durante seu tempo de vida;

- *Controle da vida útil*: Depois de todas as partículas terem suas posições atualizadas, a vida útil de cada uma delas. O PA_i do $Gbest$ é definido como zero. O restante do enxame obedece a seguinte regra

if $Fitness_i$ é pior que o $Pbest_i$ **then**

$PA_i(t) = PA_i(t - 1) + AF + WF$ **else**

$PA_i(t) = PA_i(t - 1) + AF$

end if

Se a vida útil da partícula é exaurida, ou seja, $PA_i \geq LE$, então a partícula é eliminada;

- *Gerando desafidores*: Quando o número de gerações atinge o seu valor máximo, definido anteriormente pelo usuário, um novo enxame surgirá como definido no Passo 1, e será usado para desafiar o $Gbest$;
- *Condição de parada*: Se o número de iterações é maior que o número máximo predefinido, o algoritmo para. De outra maneira, vá para o Passo 2 para uma nova iteração.

Note que o PA_i é incrementado a cada iteração pelo AF que reflete o tempo de vida útil já utilizado pela partícula, e o WF é relacionado ao *fitness* da partícula. Portanto, o problema das partículas inativas ou daquelas que podem não estar contribuindo na busca, pode ser minimizado. Desta maneira o número de partículas varia durante a busca, reduzindo o seu custo computacional (BEZERRA *et al.*, 2017).

3.3 Ferramentas para o método *online*

Espaços de busca com dimensões altas têm chamado a atenção na estimativa de funções não lineares. Uma aplicação direta dessa abordagem é a regressão, onde é feito um mapeamento não linear seguido por um processamento linear em um espaço de busca com dimensão alta. Neste trabalho são tratados problemas de regressão, onde máquinas *kernel* consideram o fato que dados históricos podem ser representados por uma combinação linear de funções *kernel* (SMOLA; VAPNIK, 1997). Métodos *kernel* têm sido aplicado com sucesso em previsão das mais diversas séries temporais (RICHARD *et al.*, 2009), velocidade do vento (KUH; MANDIC, 2009; DOWELL *et al.*, 2015), potência eólica (Kumari; Wadhvani, 2018),

demanda de potência elétrica (ESPINOZA *et al.*, 2007; CHE; WANG, 2014), entre outras.

3.3.1 Máquinas kernel

Métodos *kernel* são baseados em transformações não lineares (*kernel trick*) dos dados de entrada em um espaço de busca Hilbert (\mathcal{H}). Produtos escalares podem ser calculados usando uma função *kernel* positiva-definida ($\langle x, x \rangle \geq 0 \forall x \in X$ e vale $\langle x, x \rangle = 0 \Leftrightarrow x = 0$) satisfazendo o teorema de Mercer (VAPNIK, 2013) para produzir uma versão não linear de algoritmos de aprendizado linear convencionais.

O *kernel* Mercer é uma função contínua, simétrica, e positiva-definida $k(x_i, x_j) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, $\mathcal{X} \in \mathbb{R}^n$ ou \mathbb{C}^n , onde \mathcal{X} é um conjunto não vazio. O teorema de Mercer prova que qualquer *kernel* Mercer $k(\cdot, \cdot)$ pode ser expresso pelo produto interno de uma função não linear $\phi(x) : \mathcal{X} \rightarrow \mathcal{H}_1, x \in \mathcal{X}$,

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}_1}, \quad (3.4)$$

onde \mathcal{H}_1 é um *kernel* reproduzindo um valor \mathbb{R} ou \mathbb{C} no espaço de Hilbert, para o qual $k(\cdot, \cdot)$ é um *kernel* reproduzido, e $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ é o produto interno correspondente em \mathcal{H}_1 .

A Equação (3.4) representa um *kernel* Mercer e afirma que se x_i e x_j são mapeados em \mathcal{H}_1 por $\phi(x_i)$ e $\phi(x_j)$, respectivamente, então o produto interno dessas funções podem ser calculadas avaliando o *kernel* $k(x_i, x_j)$ mesmo se o mapeamento $\phi(\cdot)$ for desconhecido. Esse resultado é conhecido como o *kernel trick*.

Existem muitas funções *kernel*, mas a função mais comum é a função *kernel* gaussiana. Frequentemente usada em aplicações reais como nos problemas de previsão de séries temporais. Ela consiste na função de expansão para um espaço de características de dimensão infinita dado por

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / 2\sigma^2). \quad (3.5)$$

Esta função *kernel* foi adotada neste estudo. Embora existam outras opções possíveis, o *kernel* gaussiano tem uma interpretação física como uma medida de similaridade que se ajusta perfeitamente a esta aplicação em particular. Também vale a pena mencionar que superou outros candidatos, por exemplo, a função *kernel* triangular e a polinomial, conforme relatado em outros trabalhos semelhantes (ENGEL *et al.*, 2004; TOBAR *et al.*, 2014). Além disso, a escolha ou construção da função *kernel* é um problema em aberto, sendo objeto de pesquisas em andamento.

Três máquinas *kernel recursive least-squares* (KRLS) são usadas como *benchmarks*: a ALD-KRLS, a *sliding-window* KRLS (SW-KRLS), e a KRLS *tracker* (KRLS-T). Elas foram escolhidas porque a ALD-KRLS foi a primeira versão *kernel* do algoritmo adaptativo *recursive least-squares* (RLS) (VAERENBERGH; SANTAMARÍA, 2013; KABBARA; PSAROMILIGKOS, 2014). Embora não possua sistema de rastreamento, as outras duas máquinas possuem mesmo que com diferentes graus de complexidade. A derivação completa dessas máquinas de referência pode ser encontrada em Engel *et al.* (2004), Vaerenbergh *et al.* (2006), e Vaerenbergh *et al.* (2012).

Os métodos *kernel* provaram ser bem-sucedidos em aplicações onde os dados são inteiramente considerados em uma instância, ou seja, aplicativos em bateladas. No entanto, a aplicação das máquinas *kernel* para configurações *online*, onde os dados chegam sequencialmente, fornece alguns desafios já minimizados mas não resolvidos. O primeiro é o risco de *overfitting* ao usar um método no espaço de Hilbert por causa da alta dimensão dos vetores de peso. Isso tem sido tratado com o uso de regularização. Outro problema é que a complexidade da representação do estimador aumenta à medida que cresce o número de observações.

Os algoritmos KRLS calculam os coeficientes α_i , que consistem em um minimizador usado para calcular o vetor de peso ideal, resolvendo um problema de mínimos quadrados envolvendo a inversão de uma matriz *kernel* (K) cuja dimensão depende do número de exemplos armazenados (M). O segundo desafio é que a quantidade de dados processados M aumenta com o tempo em cenários *online*. Assim, os algoritmos devem restringir a quantidade de dados que serão armazenados. Como resultado, o terceiro desafio é o tempo de treinamento da batelada e/ou a atualização incremental dos algoritmos, que normalmente aumentam linearmente com o número de observações.

3.3.2 *Trabalhos relacionados*

Revisões sobre o aprendizado de múltiplos *kernels*, do inglês *multiple kernel learning* (MKL), estão disponíveis em Gönen e Alpaydın (2011) e Liu e Gu (2020). Gönen e Alpaydın (2011) concluíram que, no geral, ao usar vários *kernels* em vez de um único, se obtém melhores resultados. MKL combina um conjunto de *kernels* (*kernels* de base) de forma linear ou não, ou ainda dependente de dados em um *kernel* composto, onde os *kernels* de base podem usar diferentes funções ou valores para os hiper parâmetros de uma única função *kernel* (GÖNEN; ALPAYDIN, 2011). Inúmeros estudos têm melhorado continuamente o desenvolvimento do

MKL aplicado nas mais diversas áreas: classificação de imagens hiper espectrais (LIU; GU, 2020) e binárias (KANNAO; GUHA, 2017), previsão da qualidade do ar (ZHENG *et al.*, 2018), detecção de anomalias (GAUTAM *et al.*, 2019), categorização de objetos (WANG *et al.*, 2018), diagnóstico da doença de Alzheimer (PENG *et al.*, 2019), reconhecimento de pintores de telas em óleo (LIAO *et al.*, 2019), classificação de multiclases (WANG *et al.*, 2020), discriminando câncer em estágio inicial e final (RAHIMI; GÖNEN, 2020), agrupamentos de subespaços (REN *et al.*, 2020), entre outros.

Nos últimos anos, vários métodos combinando muitos *kernels* foram propostos. Kanna e Guha (2017) identificaram e modelaram diferentes regiões do espaço de entrada, porque cada *kernel* tem capacidade discriminativa variada em regiões distintas, nomeando-as como 'regiões de sucesso', por meio de um conjunto de funções de previsão de sucesso dependentes de instância, tendo pesos com valores altos em 'regiões de sucesso' e baixos em outras. O uso dessas funções de previsão de sucesso como funções dependentes dos pesos que promovem os *kernels* de base discriminados localmente enquanto suprime os outros. Zheng *et al.* (2018) introduziram o classificador de vetores de suporte de múltiplos *kernel*, um modelo MKL, que incorpora as características da aprendizagem por conjunto (*ensemble learning*), de *kernel* e representativa. A abordagem de alinhamento centralizado é usada para obter o peso de cada *kernel*, e uma abordagem de reforço é usada para determinar o número apropriado de *kernels*. Os *kernels* são combinados pela soma ponderada (restrição da soma cônica). O classificador de vetor de suporte é usado como a base de aprendizado e otimizado com um algoritmo de otimização genérico.

Uma abordagem MKL para a classificação de uma classe foi proposta por Gautam *et al.* (2019). O classificador usado é o SVM de uma classe de Scholkopf. O peso de cada *kernel* é definido por uma função de acionamento. O peso de cada *kernel* é atribuído localmente. Os parâmetros da função de acionamento e do classificador de uma classe são otimizados simultaneamente por meio de um processo de otimização de duas etapas. Primeiro, o problema de otimização é resolvido para encontrar os parâmetros; depois, com os parâmetros da função de acionamento atualizados, o novo peso é calculado. Wang *et al.* (2018) propuseram um algoritmo MKL dependente de dados baseado em *soft grouping*. Existem duas etapas no estágio de treinamento: (i) as amostras são divididas em grupos com um algoritmo de *soft grouping* para acomodar a correlação e diversidade das amostras; (ii) um método de otimização alternativo é usado para aprender os pesos do *kernel* e do coeficiente do vetor de suporte (classificador). O

kernel composto é determinado pelos pesos dos *kernels* dos grupos e pela probabilidade dessa amostra cair em cada grupo.

Um novo método MKL que utiliza a esparsidade estruturada definida por $l_{1,p}$ -norm ($p > 1$) como método de regularização foi projetado por Peng *et al.* (2019). Ele representa cada característica (*feature*) com um *kernel* e define a importância de cada uma através dos pesos de cada *kernel*, seguido pelo agrupamento dos *kernels* de acordo com critérios específicos da tarefa. Em seguida, uma representação ótima de *kernels* combinados de cada característica é aprendida em uma abordagem baseada nos dados. O regularizador proposto imposto aos pesos do *kernel* (l_1) é utilizado para selecionar um conjunto de características concisas dentro de cada grupo homogêneo e fundir os grupos de recursos heterogêneos tirando vantagem de normas densas (l_p). Liao *et al.* (2019) propuseram um algoritmo MKL dividido em três fases: primeiro, antes do MKL ser executado, um processo de pré-aprendizagem (K-medoids) é usado para agrupar *subkernels* candidatos semelhantes e selecionar alguns *subkernels* com melhor capacidade de classificação em cada categoria, o que diminui o tamanho dos *subkernels* candidatos; a segunda etapa calcula a capacidade de classificação de cada *kernel*, em cada categoria, para selecionar o *subkernel* com melhor desempenho de classificação; a fase final usa os *subkernels* selecionados para realizarem o MKL sob restrições l_p -norma ($p > 1$).

Um algoritmo MKL geométrico e colaborativo apresentado por Wang *et al.* (2020), classifica diretamente os dados multiclasse em classes correspondentes. Ele usa aprendizado empírico de múltiplos *kernels* para mapear a amostra em vários espaços de *kernel*, em seguida, treina a função *softmax* em cada espaço de *kernel*. A função *softmax* pode utilizar as características explícitas no espaço do *kernel* com eficiência. Para melhorar a colaboração entre diferentes espaços do *kernel*, um termo de regularização (R_U) foi projetado para exigir saídas consistentes de amostras em diferentes espaços do *kernel*. Além disso, para fazer com que as saídas das amostras tenham características de classificação geométrica, um termo de regularização de projeção geométrica (R_{G_l}) foi projetado para reduzir a distância dentro da classe das saídas das amostras em cada espaço do *kernel*. Os dois termos de regularização foram introduzidos para melhorar ainda mais a capacidade de classificação. Rahimi e Gönen (2020) formularam um método multitarefa MKL com um modelo de agrupamento em conjuntos de genes para identificar processos biológicos e aprender modelos de classificação de tarefas específicas simultaneamente. Aprendizagem multitarefa, em que diferentes tarefas são aprendidas conjuntamente, permite que tarefas companheiras com dados limitados se beneficiem de outras tarefas. Coagrupamento cria

um número predefinido de grupos de tarefas companheiras e *kernels*.

Ren *et al.* (2020) apresentaram um novo método MKL que, em conjunto, aprende um gráfico de afinidade ideal e um *kernel* de consenso adequado para fins de agrupamento. Os dados não lineares são mapeados em um espaço de Hilbert do *kernel* de reprodução em alta dimensão, onde uma análise de padrão linear é realizada. A matriz *kernel* H (matriz *kernel* Gram) é semidefinida simétrica positiva e é decomposta por meio de uma matriz quadrada auxiliar B . Esta matriz é usada para compor a matriz H com uma componente de ruído esparsa (E) para lidar com dados ruidosos. Uma estratégia de ponderação é usada como o processo de aprendizado de múltiplos *kernel*. Note que o algoritmo proposto integra o MKL com aprendizagem de estrutura local e global e a propriedade de auto-expressividade¹ do espaço de Hilbert em um modelo de otimização unificado. Um método MKL de remoção de ruídos foi apresentado por Zhou *et al.* (2019) considerando dois tipos de ruído: ruído local, que aparece em um pequeno número de elementos da matriz do *kernel* e muitas vezes é induzido por pontos fora da curva (*outliers*) ou instâncias corrompidas, e ruído global, que aparece na maioria dos elementos da matriz do *kernel* e é frequentemente induzida por *kernels* inadequados. Matrizes de ruído e tensores de ruído são introduzidos para capturar ruído local e global. Os *kernels* 'limpos' são obtidos subtraindo o ruído dos *kernels* candidatos. Para aprender o *kernel* de consenso, a discordância entre o *kernel* de consenso e todos os *kernels* 'limpos' é minimizada.

Diferentes *kernels* podem corresponder ao uso de vários conceitos de similaridade ou envolver informações provenientes de muitas fontes, ou seja, distintas representações ou subconjuntos de características. O raciocínio é semelhante a combinar diferentes classificadores. Em vez de tentar encontrar manualmente qual *kernel* funciona melhor, um algoritmo de aprendizado é responsável por selecioná-lo ou uma combinação de ambos os recursos pode ser empregada. Usar um *kernel* específico pode ser uma fonte tendenciosa e, ao permitir que o algoritmo de aprendizagem escolha entre um conjunto de *kernels*, uma solução melhorada pode ser encontrada. A função da combinação de vários *kernels* e seus parâmetros correspondentes pode ser representada como

$$k_{\eta}(x_i, x_j) = f_{\eta}(\{k_m(x_i^m, x_j^m)\}_{m=1}^P | \eta), \quad (3.6)$$

onde a função de combinação $f_{\eta} : \mathbb{R}^P \rightarrow \mathbb{R}$ pode ser linear ou não. Funções *kernel* $\{k_m : \mathbb{R}^{D_m} \times \mathbb{R}^{D_m} \rightarrow \mathbb{R}\}_{m=1}^P$ adotam (não necessariamente diferentes) P representações das características de

¹ Métodos baseados em auto-expressividade assumem que as amostras podem ser bem representadas por um combinação linear de pontos no mesmo subespaço. Normalmente, o método consiste de duas etapas: estimar uma matriz de afinidade dos dados e então aplica o agrupamento espectral para particionar os dados em grupos.

instância de dados $x_i = \{x_i^m\}_{m=1}^P$ onde $x_i^m \in \mathbb{R}^{D_m}$, e D_m é a dimensionalidade da representação da característica correspondente. η parametriza a função de combinação (GÖNEN; ALPAYDIN, 2011).

Um dos métodos mais simples para determinar a função de combinação dos *kernels* é a regra fixa. Esta estratégia usa funções sem quaisquer parâmetros, por exemplo, soma ou multiplicação dos *kernels* e não requer nenhum treinamento. Pavlidis *et al.* (2001) relataram que em uma tarefa de classificação funcional de genes, treinar uma SVM (VAPNIK, 1999) com a soma não ponderada de *kernels* heterogêneos atinge melhores resultados do que a combinação de várias SVM, cada uma treinada com um *kernel*. A abordagem heurística usa uma função de combinação parametrizada, e encontra os parâmetros dessa função geralmente observando alguma medida obtida separadamente de cada função do *kernel*. Essas medidas podem ser calculadas a partir das matrizes dos *kernels* ou tomadas como os valores de desempenho dos algoritmos de aprendizagem para cada um dos *kernels*, que são treinados separadamente. Moguerza *et al.* (2004) e Diego *et al.* (2010) propuseram uma matriz de *kernels* combinados

$$k_\eta(x_i, x_j) = \sum_{m=1}^P \eta_m(x_i, x_j) k_m(x_i^m, x_j^m), \quad (3.7)$$

onde $\eta_m(\cdot, \cdot)$ define um peso para $k_m(\cdot, \cdot)$ de acordo com x_i e x_j . Os trabalhos citados propõem diferentes heurísticas para estimar os valores da função de ponderação usando probabilidades de classes condicionais $Pr(y_i = y_j | x_i)$ e $Pr(y_j = y_i | x_j)$ calculado com uma abordagem de vizinho mais próximo. No entanto, cada função *kernel* corresponde a uma vizinhança diferente, e $\eta_m(\cdot, \cdot)$ é calculado na vizinhança induzida por $k_m(\cdot, \cdot)$.

Também é possível usar uma combinação linear em vez de uma combinação dependente de dados para formular a função *kernel* combinada da seguinte forma

$$k_\eta(x_i, x_j) = \sum_{m=1}^P \eta_m k_m(x_i^m, x_j^m), \quad (3.8)$$

onde os pesos do *kernel* são selecionados observando os valores de desempenho obtidos por cada *kernel* separadamente. Por exemplo, Qiu e Lane (2009) apresentaram duas regras simples para selecionar o peso dos *kernels* para problemas de regressão

$$\eta_m = \frac{R_m}{\sum_{h=1}^P R_h} \quad (3.9)$$

e

$$\eta_m = \frac{\sum_{h=1}^P M_h - M_m}{(P-1) \sum_{h=1}^P M_h}, \quad (3.10)$$

onde R_m é o coeficiente de correlação de Pearson entre as saídas verdadeiras e as previstas geradas pelo regressor usando a matriz do *kernel* K_m , e M_m é o erro médio quadrático gerado pelo regressor usando a matriz do *kernel* K_m . Essas heurísticas encontram uma combinação convexa dos *kernels* de entrada como o *kernel* combinado.

O coeficiente de correlação de Pearson é uma medida comum de associação entre duas variáveis contínuas. É definido como a razão da covariância das duas variáveis e o produto de seus desvios-padrão, comumente denotado pela letra grega ρ , onde

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (3.11)$$

O coeficiente de correlação da amostra, R , pode ser obtido inserindo a covariância e os desvios padrão da amostra na fórmula anterior, ou seja,

$$R = \frac{\sum_{i=1}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.12)$$

O coeficiente de correlação de Pearson varia de -1 a 1 . Quando $\rho > 0$, duas variáveis tendem a aumentar ou diminuir simultaneamente; para $\rho < 0$, uma variável tende a aumentar quando a outra diminui; finalmente, $\rho = 0$ corresponde à ausência de associação (SHESKIN, 2000).

Existem duas diferenças principais entre os modelos descritos anteriormente e o apresentado neste trabalho. Em primeiro lugar, o algoritmo de aprendizado dos modelos anteriores é uma SVM, enquanto o proposto usa algoritmos KRLS, que produzem soluções muito mais esparsas e com maior robustez ao ruído. Além disso, as máquinas KRLS são totalmente *online* projetados para operar em ambientes de tempo real, onde os dados ficam disponíveis uma amostra por vez. Em segundo lugar, quase todos os modelos anteriores usam uma combinação linear, que é a abordagem mais popular com duas categorias básicas: soma não ponderada, ou seja, usando a soma ou média dos *kernels* como o *kernel* combinado; e a soma ponderada, que neste caso a seguinte função de combinação pode ser parametrizada linearmente

$$k_\eta(x_i, x_j) = f_\eta(\{k_m(x_i^m, x_j^m)\}_{m=1}^P | \eta) = \sum_{m=1}^P \eta_m k_m(x_i^m, x_j^m), \quad (3.13)$$

onde η denota os pesos dos *kernels*. Outras versões desta abordagem diferem nas restrições: a soma linear ($\eta \in \mathbb{R}^P$); a soma cônica ($\eta \in \mathbb{R}_+^P$); ou a soma convexa ($\eta \in \mathbb{R}_+^P$ e $\sum_{m=1}^P \eta_m = 1$). Os autores em (DIEGO *et al.*, 2010) aplicaram uma combinação não linear que usa funções não lineares de *kernels*, por exemplo, multiplicação, potência e exponenciação. O modelo apresentado usa métodos de combinação dependentes de dados que atribuem pesos específicos

aos *kernels* para cada instância de dados. Fazendo isso, é possível identificar distribuições locais nos dados e aprender as regras de combinação dos *kernels* adequadas para cada região.

3.4 *Multiple challengers kernels recursive least-squares (MC-KRLS)*

O principal objetivo deste estudo é apresentar um modelo de máquina de aprendizagem *multi kernel* que possa lidar com problemas de regressão com um dicionário menor e não estático, e obter resultados melhores ou semelhantes quando comparados com suas respectivas versões *single kernel*.

O procedimento de ajuste do modelo é totalmente orientado por dados, tornando-o ideal para aplicações de *smart grid* onde muitos geradores compartilham um sistema de energia altamente interconectado e o uso de dependência espacial é desejável. Os resultados obtidos com diferentes máquinas *kernel* são comparados em dois cenários, primeiro considerando apenas o aspecto temporal do conjunto de dados, enquanto a dependência espacial é analisada posteriormente.

O ALD-KRLS pode resolver as dificuldades apresentadas nas máquinas *kernel*, o crescimento da matriz *kernel* e o problema de *overfitting*. Ao aplicar um procedimento de esparsificação à matriz *kernel*, pode-se limitar o tamanho do dicionário e evitar o *overfitting*, mas eventualmente o dicionário atingirá seu tamanho máximo e não aprenderá com as novas informações recebidas pelo modelo.

O método proposto, denominado MC-KRLS, é elaborado para usar múltiplos algoritmos ALD-KRLS ou outras máquinas *kernel* quaisquer, adotando a mesma entrada, mas com dicionários diferentes que se relacionam entre si pelo seu tamanho. Assim, é sempre possível aprender à medida que novas informações são recebidas, e controlar o tamanho dos dicionários nas máquinas *kernel*. Para chegar a esse resultado, os esforços de pesquisa se concentraram em quando criar uma nova matriz *kernel*, como calcular a previsão combinada, e quando excluir a matriz *kernel* (BEZERRA *et al.*, 2021).

3.4.1 *Quando criar uma nova matriz do kernel*

O método adotado estabelece uma dependência entre as matrizes dos *kernels*, permitindo que uma nova matriz seja criada quando a anterior atingir um tamanho pré-determinado definido pelo usuário. Por exemplo, vamos considerar um MC-KRLS com três *kernels* (M3C-

KRLS) (BEZERRA *et al.*, 2021).

O primeiro *kernel* é criado como no algoritmo ALD-KRLS, mas o segundo *kernel* é gerado somente depois que o primeiro atinge 25% do tamanho máximo de sua matriz. O terceiro é iniciado apenas quando o segundo atinge 15% do tamanho máximo de sua matriz. Note que os percentuais aqui citados são exemplos, ou seja, o usuário pode definir qualquer outra proporção.

3.4.2 Como calcular a previsão combinada (\hat{Y})

O primeiro método usado é a média aritmética de cada saída da matriz dos *kernels*. Este é considerado um método padrão e nenhum acrônimo é atribuído a ele. O segundo método é uma média ponderada de cada saída das matrizes dos *kernels* conforme descrito na Equação (3.14). A versão ponderada é indicada por MC-KRLSW (BEZERRA *et al.*, 2021).

$$\hat{Y} = \frac{w_{k_1}\hat{Y}_{k_1} + w_{k_2}\hat{Y}_{k_2} + \dots + w_{k_n}\hat{Y}_{k_n}}{w_{k_1} + w_{k_2} + \dots + w_{k_n}}, \quad (3.14)$$

onde w_{k_n} é o peso associado com a saída do *kernel* n , que é calculado da seguinte forma

$$w_{k_i} = \frac{1}{E_i} / \sum_{j=1}^n \frac{1}{E_j}, \quad (3.15)$$

onde E_i é o erro absoluto, *absolute error* (AE), da previsão anterior da matriz do *kernel* i

$$E_i = |Y - \hat{Y}_{k_i}|. \quad (3.16)$$

3.4.3 Quando apagar a matriz de um kernel

Um método adaptativo para lidar com as máquinas *kernel* e escolher as ‘melhores’ é descrito nesta seção. A primeira abordagem assume que quando todas as matrizes dos *kernels* atingirem seu tamanho máximo, um contador será iniciado e quando atingir o seu valor máximo o erro médio quadrático, *mean square error* (MSE), será calculado. A matriz do *kernel* com o pior MSE é então excluída. Este é o método padrão e nenhum acrônimo é fornecido a ele (BEZERRA *et al.*, 2021).

O segundo método é semelhante ao primeiro, mas em vez de calcular o MSE quando o contador termina, o AE é calculado para cada iteração. Assim, a melhor matriz *kernel*, ou seja, com o menor AE, não sofrerá nenhuma alteração, enquanto as demais matrizes terão suas ‘idades’ incrementadas. A matriz que atingir primeiro a idade máxima predefinida é excluída (BEZERRA *et al.*, 2021).

A LE é uma medida estatística de quanto tempo um organismo pode viver, e em uma determinada idade espera-se que a vida cesse. O LE é baseado em muitas características, como o ano de nascimento, idade atual, e outros fatores demográficos. Uma versão simplificada desse conceito é usada aqui com base em duas constantes a AF e WF.

Na natureza, os líderes da colônia são constantemente desafiados por novos indivíduos. O envelhecimento facilita a substituição de um líder por um indivíduo mais jovem, o que provavelmente criará mais oportunidades de diversidade e melhorias. Inspirado neste fenômeno, este trabalho adapta a ideia acima citada da natureza para as máquinas *kernel* e propõe o MC-KRLS agregando AF e WF, resultando no algoritmo MC-KRLSA.

O envelhecimento não é a única característica que compõe o LE de um indivíduo. Neste estudo, WF representa os outros recursos associados ao LE. WF desempenha um papel sempre que o *kernel* (K_i) não consegue encontrar um novo resultado ($\hat{y}_{k_i}(t)$) melhor do que o resultado anterior ($\hat{y}_{k_i}(t-1)$). Assim, mesmo *kernels* da mesma idade terão LE diferentes e deixarão de existir em momentos distintos da busca. Primeiro, a idade do *kernel* KA_i do K_i é definida como zero, e então as máquinas *kernel* seguem as regras definidas na Seção 3.2.

O algoritmo proposto visa superar as principais limitações de abordagens semelhantes no rastreamento de mudanças no processo estocástico subjacente, conforme discutido na Seção 3.4. Além do número de *kernels* usados (η_K), seus parâmetros são os seguintes: o valor máximo atingido pelo contador (*counter.max*), que neste trabalho é definido com o mesmo valor do tamanho máximo do dicionário; e quando criar uma nova matriz, que é investigada para 25%, 50%, 75%, e 100% do tamanho máximo do dicionário (M). Para fins de reprodutibilidade destes modelos, os códigos-fonte estão disponíveis para *download* em Bezerra (2020).

A Tabela 3 mostra as siglas utilizadas que são uma composição de MC-KRLS, o número de *kernels* utilizados, como a previsão é combinada, e o método de exclusão.

Tabela 3 – Modelos MC-KRLS utilizados

<i>Kernels</i>	Previsão Combinada	Método de Exclusão	Acrônimo
2	Média Aritmética	Contador	M2C-KRLS
3	Média Aritmética	Contador	M3C-KRLS
2	Média Ponderada (W)	Contador	M2C-KRLSW
3	Média Ponderada (W)	Contador	M3C-KRLSW
2	Média Aritmética	Envelhecimento (A)	M2C-KRLSA
3	Média Aritmética	Envelhecimento (A)	M3C-KRLSA
2	Média Ponderada (W)	Envelhecimento (A)	M2C-KRLSAW
3	Média Ponderada (W)	Envelhecimento (A)	M3C-KRLSAW

Fonte: Bezerra *et al.* (2021).

3.5 Conclusões

Neste capítulo foram apresentados conceitos sobre ANN, BP, PSO, e algumas outras ferramentas de IC utilizadas durante a pesquisa. Foram apontadas dificuldades enfrentadas pelo BP, justificando o interesse na pesquisa do PSO como algoritmo de aprendizagem, apresentando também suas limitações, junto com uma revisão de algumas modificações do PSO sugeridas por diversos autores. Em seguida foi proposta a introdução de dois fatores, o WF e o AF no PSO que busca minimizar as limitações encontradas no PSO clássico.

No restante do capítulo, o conceito de métodos *kernel* foi apresentado. Apontado os dois problemas mais comuns na utilização dos métodos *kernel* em ambientes *online*, especificamente, problemas de *overfitting* e de custo computacional, junto com uma série de soluções. Também foram discutidos alguns métodos *kernel*. Neste trabalho é proposto um algoritmo que ameniza os dois problemas, relacionando as máquinas *kernels* de maneira adaptativa através do tamanho de seus dicionários e suas performances utilizando o AF e o WF.

4 ESTUDO DE CASO

Neste capítulo são apresentados os respectivos estudos da aplicação dos fatores de envelhecimento e enfraquecimento no processo de aprendizagem de ferramentas¹ *offline* (Seção 4.1) e *online* (Seção 4.2).

4.1 Método *offline*

Nesta seção, o comportamento de busca do PSO-AWF resolvendo problemas de regressão é estudado através de experimentos. Os efeitos do controlador do tempo de vida são analisados. Em particular, respostas para duas perguntas são procuradas: Como o PSO-AWF funciona em problemas de regressão, e qual a melhor maneira de definir os parâmetros do modelo FTDNN. As Figuras 6, 7 e 8 apresentam diagramas de blocos que descrevem os processos de seus respectivos estudos de caso, os paralelogramos indicam os conjuntos de dados utilizados (Seção 4.1.1), os triângulos como os dados foram extraídos e divididos (Seção 4.1.2), os retângulos as ferramentas utilizadas no processo de treinamento (Seção 3.1), e por fim o porquê dos diferentes arranjos (Seção 4.1.3).

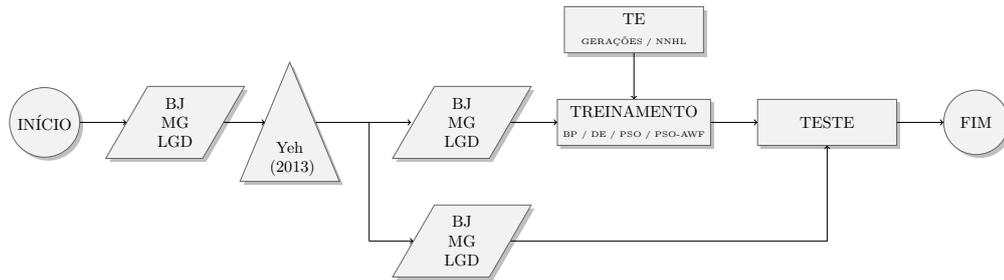
4.1.1 Conjuntos de dados

Esta seção apresenta os conjuntos de dados utilizados nos experimentos. O conjunto de dados de vento, *wind dataset* (WS), e três conjuntos de dados *benchmarks* chamados: *Box-Jenkins gas furnace data* (BJ) (BOX; JENKINS, 1970), *Mackey-Glass chaotic time series* (MG) (MACKEY; GLASS, 1977), e *Data set A: laser generated data* (LGD) (WEIGEND; GERSHENFELD, 1994). Eles foram utilizados para testar o desempenho do algoritmo proposto, o PSO-AWF. Estas séries temporais *benchmarks* são amplamente utilizadas como conjuntos de dados para testar o desempenho de ANN (HUEBNER *et al.*, 1989; GERSHENFELD; WEIGEND, 1993; WEIGEND; GERSHENFELD, 1994).

O conjunto de dados BJ foi gravado a partir do processo de combustão de uma mistura de ar e metano (BOX; JENKINS, 1970; HYNDMAN, 2021). Os dados originais incluem 296 amostras de concentração de CO_2 como variável de saída $y(t)$, e o fluxo de entrada de gás como variável de entrada $u(t)$. A quantidade de dados utilizados durante o processo de aprendizagem

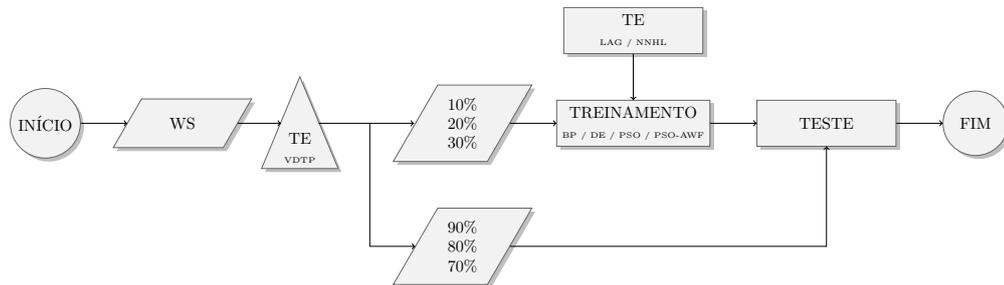
¹ Note que os termos ferramentas *offline* e *online* indicam como e quando os parâmetros dos modelos são definidos conforme o Capítulo 3.

Figura 6 – Diagrama de blocos dos processos do estudo de caso 1.



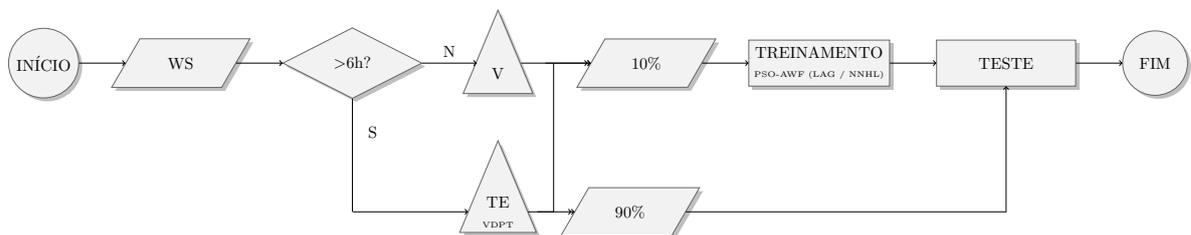
Fonte: Elaborado pelo autor (2021).

Figura 7 – Diagrama de blocos dos processos do estudo de caso 2.



Fonte: Elaborado pelo autor (2021).

Figura 8 – Diagrama de blocos dos processos do estudo de caso 3.



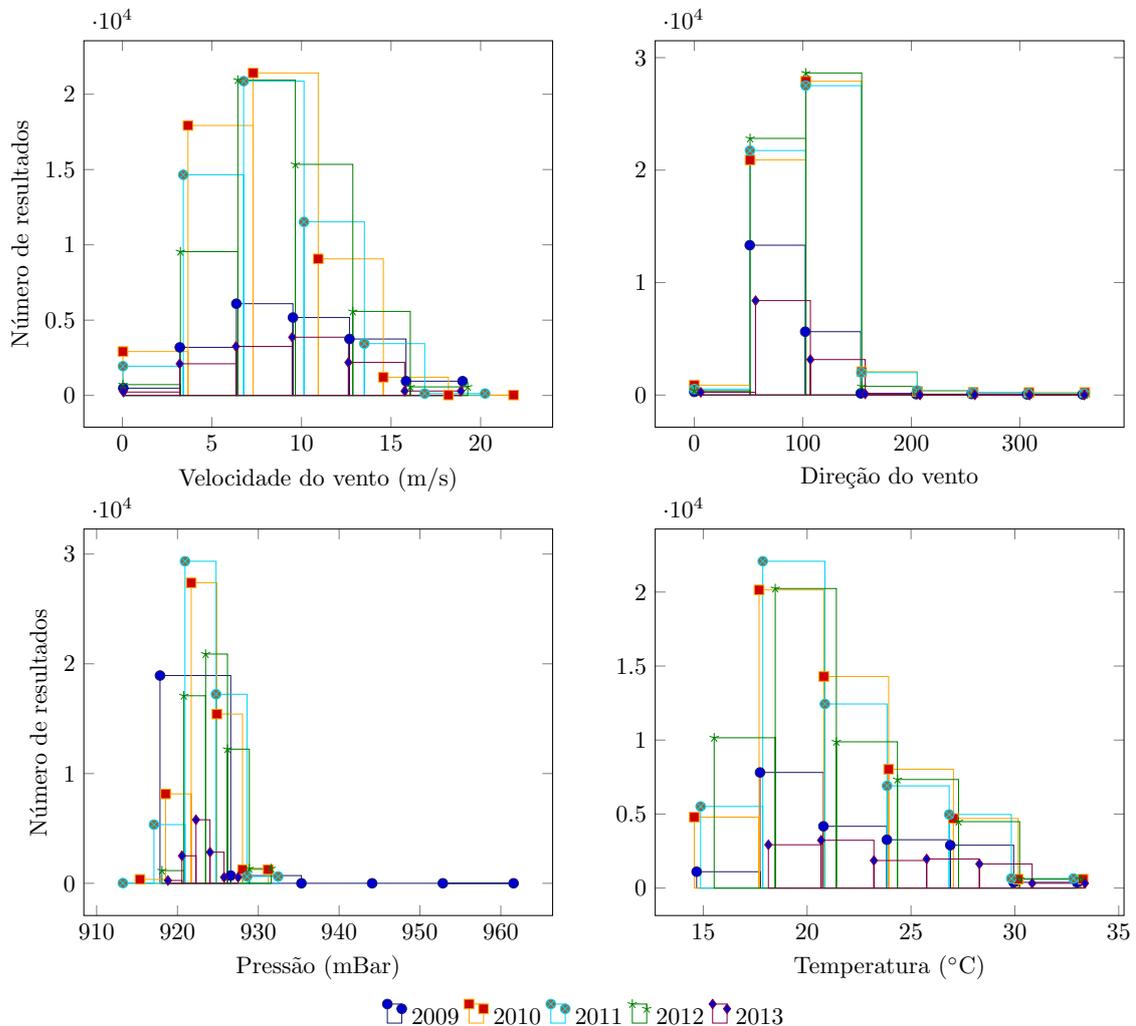
Fonte: Elaborado pelo autor (2021).

da FTDNN (dados de treinamento) foi de 140 amostras, e outras 150 foram definidas como um conjunto de testes utilizados para verificar a eficácia do treinamento. O conjunto de dados MG é gerado a partir da equação diferencial ordinária de *Mackey-Glass* (MACKEY; GLASS, 1977). Os dados de treino consistem em 450 amostras, e 500 foram utilizadas como conjunto de teste. O conjunto de dados LGD é um dos seis conjuntos de dados preparado para o *Santa Fe Time Series Competition* (WEIGEND; GERSHENFELD, 1994). Uma descrição detalhada das medições pode ser encontrada em Huebner *et al.* (1989). Os dados de treinamento consistem em 400 amostras, e outras 400 como conjunto de teste. Para os experimentos executados nesta pesquisa, o número de amostras utilizadas para as fases de aprendizagem e teste foram escolhidos com base em Yeh (2013), bem como outros parâmetros que serão apresentados na próxima seção.

As medições do conjunto WS foram feitas a uma altura de 100 metros. A torre é localizada no estado de Pernambuco, Brasil. Os dados disponíveis são: médias da velocidade do

vento (v), direção do vento (D), pressão (p), e temperatura (T), registrados a cada 10 minutos a partir de 17/08/2009 até 24/03/2013. Diferente dos conjuntos de dados *benchmarks* que estão livremente disponíveis para uso e são bem conhecidos, o conjunto de dados WS é fornecido por uma empresa privada, e por esta razão apresentamos a Figura 9 onde 4 histogramas, um para cada variável do WS, são utilizados para extrair mais informações sobre o WS. Como pode ser visto, não há dados inconsistentes, tais como valores negativos ou extremos, e independente do ano, os valores de mínimo e máximo são próximos, bem como a sua distribuição. As turbinas eólicas geralmente começam a gerar com velocidades do vento entre 3 e 4 m/s, e no local avaliado, aproximadamente 95% do tempo é favorável para a produção de energia eólica.

Figura 9 – Histograma do conjunto de dados WS entre 2009 e 2013.



Fonte: Elaborado pelo autor (2021).

O uso de rosa dos ventos (Apêndice A) mostra que os ventos leste e nordeste são predominantes, pequenas diferenças ocorrem nos anos de 2009 e 2013 por não existirem dados

completos destes anos. O mesmo ocorre nas análises mês à mês, espaços em branco foram propositalmente criados para indicar a falta do conjunto de dados.

4.1.2 Definições de parâmetros

Todos os conjuntos de dados de referência foram utilizados como entrada de uma FTDNN para prever o próximo passo, com uma entrada composta por: $bias$, $x(t)$, $x(t-1)$, $x(t-2)$, e $x(t-3)$. O conjunto de dados WS foi usado como entrada de uma FTDNN para prever 30 minutos, 1, 3 e 6 horas à frente, com uma entrada composta por: $bias$, $x(t), \dots, x(t-LAG)$. O número de LAG foi definido como 1, 3, 5 ou 7 para o método de tentativa e erro (TE) e $LAG \in \mathbb{Z} \mid 1 \leq LAG \leq 7$ quando o PSO-AWF foi utilizado para definir estes parâmetros. Todos os conjuntos de dados foram normalizados entre 0,1 e 0,9 por causa da função de ativação adotada, uma sigmoide logística.

Para os conjuntos de dados BJ, MG e LGD, foram utilizados os mesmos valores para os parâmetros: número de gerações (250, 500, 750 ou 1000) e o número de neurônios na camada oculta, do inglês *number of neurons in the hidden layer* (NNHL) (1, 2, 3, 4, ou 5). Cada combinação entre o número de gerações e NNHL foi repetido 50 vezes, gerando um total de 12000 FTDNN, 3000 FTDNN para cada método de treinamento. O PSO e PSO-AWF utilizaram o mesmo número de partículas (120). Os parâmetros do PSO-AWF: LE igual a 150, AF igual a 1, e WF igual a 5. Uma nova nuvem com 120 partículas foi gerada a cada 150 iterações.

Além da divisão dos conjuntos de dados BJ, MG e LGD, com base em Yeh (2013), uma busca dos melhores parâmetros foi conduzida através do método de validação cruzada *k-fold* (JAMES *et al.*, 2013) com 5 blocos de medições de tamanhos iguais. Um desses blocos foi deixado de fora do processo de aprendizado e utilizado como conjunto de validação, e o conjunto de teste foi do mesmo tamanho utilizado anteriormente, 150, 500 e 400 medições, repetidos 50 vezes cada um. Para ser justo, as mesmas partições dos dados de treinamento, validação e teste são usadas quando diferentes algoritmos são treinados no conjunto de dados. Observa-se que não existe uma regra de divisão de dados, apenas práticas comuns. O erro médio quadrático, MSE, é usado como o índice de desempenho.

Quanto ao conjunto de dados WS, o número de gerações foi definido como 25. O NNHL foi definido como 3, 5, 7 ou 11 usando TE, e $NNHL \in \mathbb{Z} \mid 3 \leq NNHL \leq 11$ usando o PSO-AWF. Cada combinação entre o número de gerações e NNHL foi repetido vinte vezes com BP e duas vezes usando PSO e PSO-AWF, gerando um total de 39936 FTDNN, onde 30720

(BP), 3072 (DE), 3072 (PSO) e 3072 (PSO-AWF). O número de partículas foi o mesmo (110) para o PSO e PSO-AWF. Os parâmetros do PSO-AWF: LE igual a 10, AF igual a 1, e WF igual a 2. Uma nova nuvem com 60 partículas foi gerada a cada 10 iterações.

Para o segundo estudo de caso, o conjunto de dados WS foi dividido em 10%, 20% e 30% (aproximadamente um ano) de dados para treinamento e o restante para validação. A divisão foi definida de acordo com o interesse da investigação, verificar diferentes arranjos para se obter resultados semelhantes ou melhores em um intervalo de tempo menor. Por isso a decisão de diminuir o tamanho do conjunto de treinamento de 30% para 20% e 10%.

Para todos os conjuntos de dados, a taxa de aprendizagem do BP foi 0,7. Os pesos sinápticos iniciais foram valores aleatórios entre -10 e 10. Os pesos de inércia do PSO e PSO-AWF foram linearmente reduzidos de 0,9 a 0,5, $C_1 = C_2 = 1$, e as velocidades iniciais foram zero. DE/rand/1/bin com $F = 0.5(1 + rand())$, o CR aumentou linearmente entre 0,5 a 1,0 e o número de partículas foi o mesmo usado pelo PSO e PSO-AWF.

O intervalo de normalização, o número de gerações, a taxa de aprendizagem do BP, e as velocidades iniciais foram escolhidos com base em Yeh (2013), enquanto os outros parâmetros foram aleatórios, sob a suposição de que o usuário não tem prévio conhecimento para definir estes parâmetros, sendo assim, definidos durante a busca dos melhores resultados.

4.1.3 Resultados

Três experimentos foram feitos com estas séries temporais, identificados como

- Estudo de caso 1, *test case 1* (TC1), testa todos os métodos de aprendizado (BP, DE, PSO e PSO-AWF) com os três bancos de dados *benchmarks* usando FTDNN com diferentes números de gerações (250, 500, 750 ou 1000) e NNHL (1, 2, 3, 4, ou 5) escolhidos através da TE;
- Estudo de caso 2, *test case 2* (TC2), testa todos os métodos de aprendizado com o banco de dados WS usando FTDNN com diferentes números de NNHL (3, 5, 7 ou 11) e LAG (1, 3, 5 ou 7) escolhidos por TE; e
- Estudo de caso 3, *test case 3* (TC3), testa o algoritmo PSO-AWF com o banco de dados WS usando FTDNN com diferentes números de NNHL $\in \mathbb{Z} \mid 3 \leq \text{NNHL} \leq 11$, e LAG $\in \mathbb{Z} \mid 1 \leq \text{LAG} \leq 7$ escolhidos pelo PSO-AWF.

Os resultados obtidos no TC1 foram utilizados para verificar se o PSO-AWF produz um menor custo computacional, e comportamento semelhante ao PSO. Os resultados do TC2

foram usados para determinar a melhor estrutura da ANN usando o banco de dados WS, e finalmente, os resultados do TC3 foram usados para verificar se existe redução do custo computacional quando o PSO-AWF é usado para definir os parâmetros NNHL e LAG, além de realizar o treinamento da ANN, quando comparado aos resultados obtidos pelo TC2.

Todos os resultados experimentais foram categorizados através do MSE da fase de teste (representado por MSE), e o custo computacional (representado por t) em segundos, além de estatísticas em cada categoria que incluem a média, mediana e desvio padrão das estimativas, e o MSE, representados pelo diagrama de caixas e Taylor, respectivamente.

4.1.3.1 Estudo de caso 1

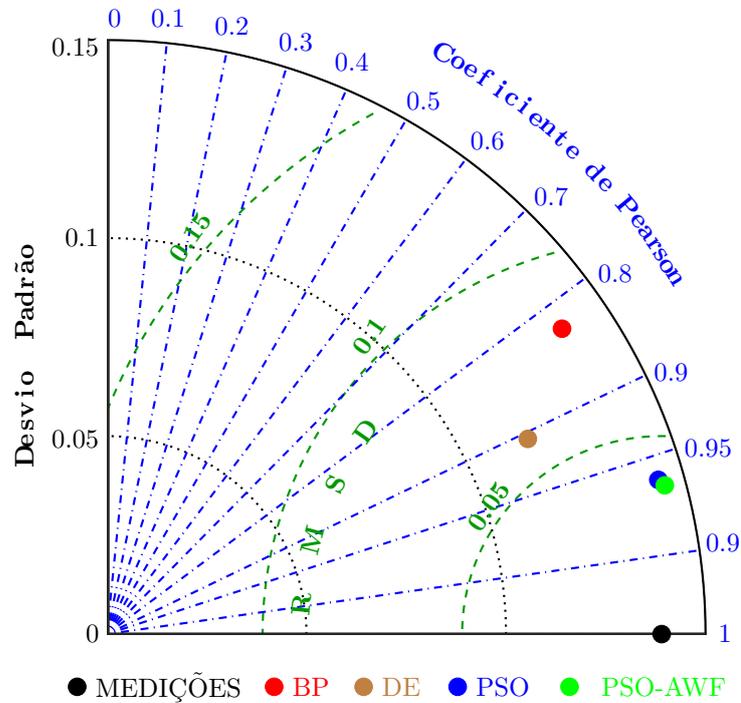
A avaliação detalhada do conjunto de dados LGD é apresentada a seguir. O resultado desse conjunto de dados e dos outros será condensando na Tabela 4. O diagrama de Taylor (TAYLOR, 2001) fornece uma representação gráfica que resume o quão perto o padrão estimado está dos dados observados, quantificados através de seus coeficientes de correlação de Pearson, RMSE e a amplitude de suas variações representadas por seus desvios padrões.

A Figura 10 é um diagrama de Taylor, que mostra a capacidade com que os modelos de regressão estimam a série temporal LGD. Estatísticas para quatro modelos foram computadas, e um ponto colorido foi atribuído a cada modelo. A posição de cada ponto no gráfico quantifica quão perto os dados simulados por este modelos estão das medições (ponto preto). Considere o ponto vermelho (modelo BP), por exemplo, seu coeficiente de correlação com as observações é de cerca de 0,83 (linhas radiais azuis) sendo considerado uma baixa correlação, resultando em um RMSE relativamente alto. Os contornos verdes indicam os valores do *root mean square deviation* (RMSD) e pode ser visto que, no caso do modelo BP é cerca de 0,08. O desvio padrão do padrão dos dados simulados é proporcional à distância radial a partir da origem (linha pontilhada preta). Para o modelo BP o desvio padrão dos dados simulado é maior do que o desvio padrão observado.

Os méritos relativos a vários modelos podem ser inferidos a partir da Figura 10. Padrões simulados que estão próximos aos padrões observados serão posicionados próximos do ponto preto localizado no eixo x. Esses modelos terão correlação relativamente alta e baixo RMSE. Modelos que são posicionados no mesmo arco imaginário preto pontilhado das medições terá o desvio padrão correto (o que indica que as variações de padrão têm a mesma amplitude). Na Figura 10 pode ser visto que os modelos PSO e PSO-AWF geralmente têm resultados que se

aproximam das medições, cada um com aproximadamente o mesmo RMSE e correlação com as observações. No entanto, ambos os modelos têm um desvio padrão ligeiramente maior do que o observado.

Figura 10 – Diagrama de Taylor para 4 modelos (LGD).



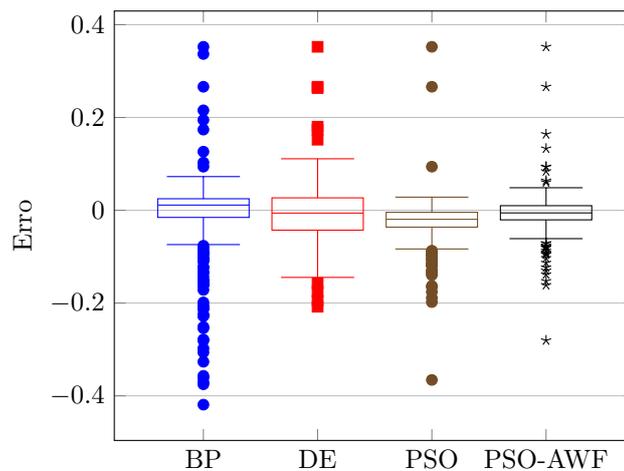
Fonte: Bezerra *et al.* (2017).

Agora que se tem uma visão geral de cada modelo, será utilizado o diagrama de caixas (TUKEY, 1977). Método padronizado para exibir a distribuição de dados com base nas seguintes informações: limite inferior, primeiro quartil, mediana, terceiro quartil e limite superior. No gráfico de caixa mais simples, o retângulo central, abrange o primeiro quartil até o terceiro quartil chamado intervalo interquartil, *interquartile range* (IQR). Um segmento dentro do retângulo mostra a mediana e retas acima e abaixo da caixa mostram os locais dos limites mínimo e máximo.

A Figura 11 é um diagrama de caixas que resume algumas características da distribuição produzido por cada modelo de regressão que estima o próximo passo da série histórica LGD. O BP tem uma grande quantidade de valores discrepantes, *outliers*, apresentados através de pontos azuis após os limites inferior e superior. As quatro seções do gráfico de caixa são desiguais em tamanho. Isto mostra que o modelo tem estimativas semelhantes em determinadas partes da escala, mas em outras partes há uma variabilidade elevada nas estimativas. Os modelos DE têm os quartis mais largos. Os modelos que utilizam o PSO e PSO-AWF têm distribuições

similares com medianas próximas (linhas marrom e preta). Observando os limites inferior e superior do PSO percebe-se que têm limite inferior mais longos que o superior, isso significa que as estimativas são mais variadas entre o grupo quartil mais negativo, já o PSO-AWF possui a mediana mais próxima de zero e os menores quartis. Em comparação com os valores discrepantes inferiores, os superiores são mais espalhados, o que significa que as estimativas variam ao longo do grupo de quartil mais positivo e muito semelhantes no grupo de quartil menos positivo.

Figura 11 – Diagrama de caixas do próximo passo (LGD).



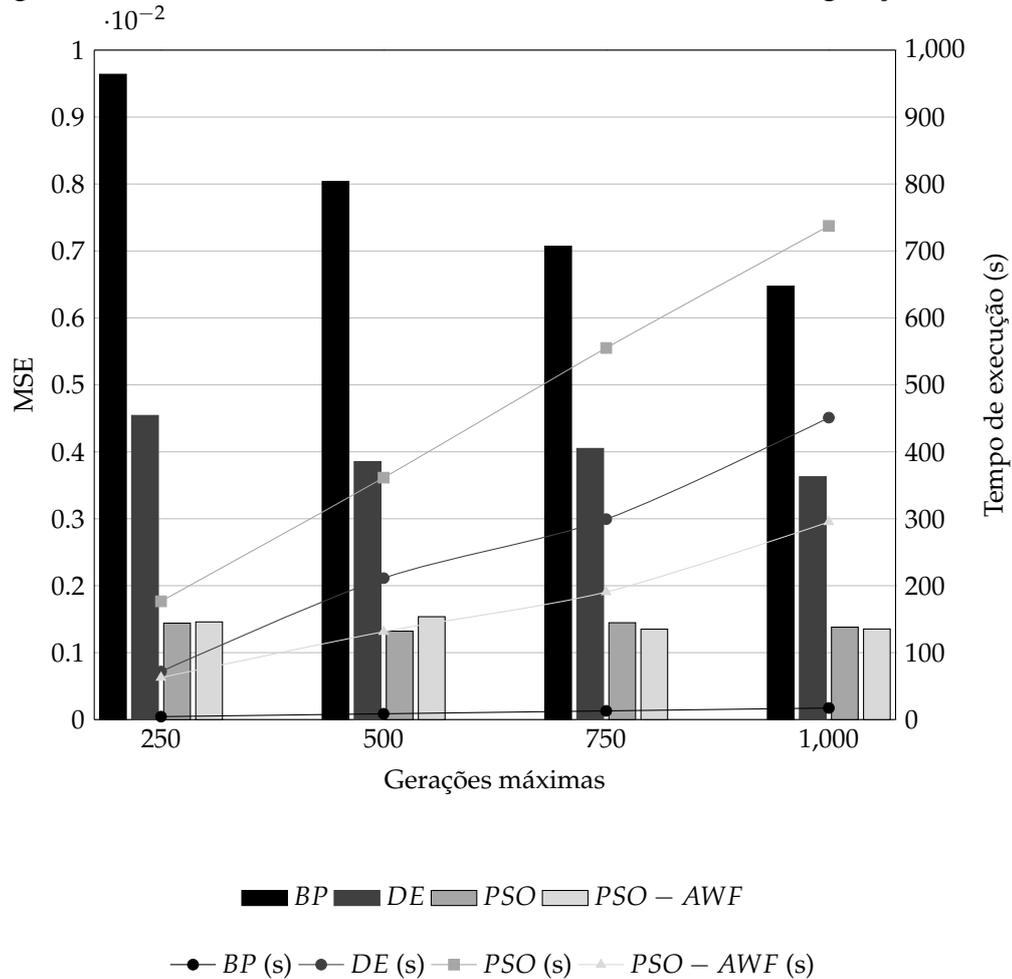
Fonte: Bezerra *et al.* (2017).

A Figura 12 é um diagrama que resume a relação entre o valor máximo de gerações, o MSE (barras), e o tempo de execução (linhas) com os quais modelos de regressão estimam o próximo passo da série histórica LGD. Como esperado, o tempo de execução é diretamente proporcional ao número de gerações, à medida que aumenta o número de gerações (250, 500, 750, 1000), o tempo de execução também aumenta. O mesmo comportamento foi encontrado nos três conjuntos de dados *benchmarks*. O DE, PSO, e PSO-AWF não apresentaram qualquer tipo de tendência entre o MSE e o número de gerações. No entanto, o BP diminuiu o valor do MSE com o aumento do valor de gerações máximas.

Para evitar inúmeras repetições das Figuras 10, 11 e 12, a Tabela 4 resume os resultados calculados pelos quatro métodos de treinamento (BP, DE, PSO e PSO-AWF), utilizando uma FTDNN para prever os dados dos conjuntos BJ, MG e LGD. Na tabela encontramos o NNHL, o tempo de execução e o MSE para cada método de treinamento em cada conjunto de dados, sendo repetido com diferentes números de gerações.

Em geral, o PSO-AWF supera o algoritmo PSO no tempo de execução com um valor de MSE aproximado para prever o BJ, MG, e LGD com o FTDNN. O BP tem um melhor tempo

Figura 12 – Performance do FTDNN sobre diferente números de geração máxima.



Fonte: Bezerra *et al.* (2017).

de execução em todos os conjuntos de dados; no entanto, é menos eficiente do que os outros algoritmos. O BP realiza pior previsões em todos os bancos de dados. O DE tem um tempo de execução variando entre os tempos do PSO e PSO-AWF; no entanto, é menos eficiente.

Tabela 4 – Resultados da FTDNN com diferente valores de NNHL.

Dados	Método	250.º Geração			500.º Geração			750.º Geração			1000.º Geração		
		NNHL	t (s)	MSE (10^{-5})	NNHL	t (s)	MSE (10^{-5})	NNHL	t (s)	MSE (10^{-5})	NNHL	t (s)	MSE (10^{-5})
BJ	BP	5	4	641	5	3	645	5	4	641	5	6	638
	DE	1	116	630	5	273	678	1	339	617	3	540	608
	PSO	5	246	632	4	486	623	5	740	628	3	973	624
	PSO-AWF	2	64	596	5	137	624	1	164	610	5	251	632
MG	BP	5	5	264	5	10	256	3	14	247	3	19	235
	DE	3	81	526	3	264	363	2	314	338	3	438	332
	PSO	3	304	219	5	416	220	4	615	218	4	809	217
	PSO-AWF	4	61	227	5	118	220	5	156	221	4	221	220
LGD	BP	5	4	963	5	9	803	5	13	707	5	17	647
	DE	2	71	454	2	211	385	5	299	405	2	451	362
	PSO	3	177	144	5	361	132	5	555	144	5	737	138
	PSO-AWF	5	63	145	3	131	153	4	190	135	5	295	135

Fonte: Bezerra *et al.* (2017).

A Tabela 5 resume os resultados calculados pelos quatro métodos de treinamento

(BP, DE, PSO e PSO-AWF), utilizando uma FTDNN para prever os dados dos conjuntos BJ, MG e LGD. Na tabela encontramos o NNHL, o tempo de execução e o MSE para cada método de treinamento em cada conjunto de dados, sendo repetido com diferentes números de gerações. Desta vez a parametrização é obtida pelo método validação cruzada *k-fold* com 5 blocos de tamanhos iguais. Um desses blocos foi deixado de fora do processo de aprendizado e utilizado como conjunto de validação, e o restante dos dados foram usados como conjunto de teste (JAMES *et al.*, 2013).

A aplicação do PSO-AWF supera novamente os outros algoritmos no banco de dados BJ, para o MG e LGD o PSO-AWF tem o melhor resultado que antes era do PSO. O comportamento do BP e do DE foi semelhante ao encontrado na Tabela 4.

Tabela 5 – Resultados da FTDNN com diferente valores de NNHL (5-fold).

Dados	Método	250. ^o Geração		500. ^o Geração		750. ^o Geração		1000. ^o Geração	
		NNHL	MSE (10^{-5})	NNHL	MSE (10^{-5})	NNHL	MSE (10^{-5})	NNHL	MSE (10^{-5})
BJ	BP	5	781	5	765	5	756	5	744
	DE	3	767	5	782	5	803	3	708
	PSO	4	721	1	727	2	726	2	720
	PSO-AWF	1	673	3	680	1	681	3	683
MG	BP	3	277	4	273	5	264	5	265
	DE	4	270	4	265	2	277	2	269
	PSO	3	264	4	266	5	261	5	263
	PSO-AWF	5	265	4	260	4	259	5	271
LGD	BP	5	1054	5	977	4	856	5	791
	DE	3	513	5	476	2	466	2	547
	PSO	5	169	5	171	5	161	3	164
	PSO-AWF	3	166	4	169	5	151	5	178

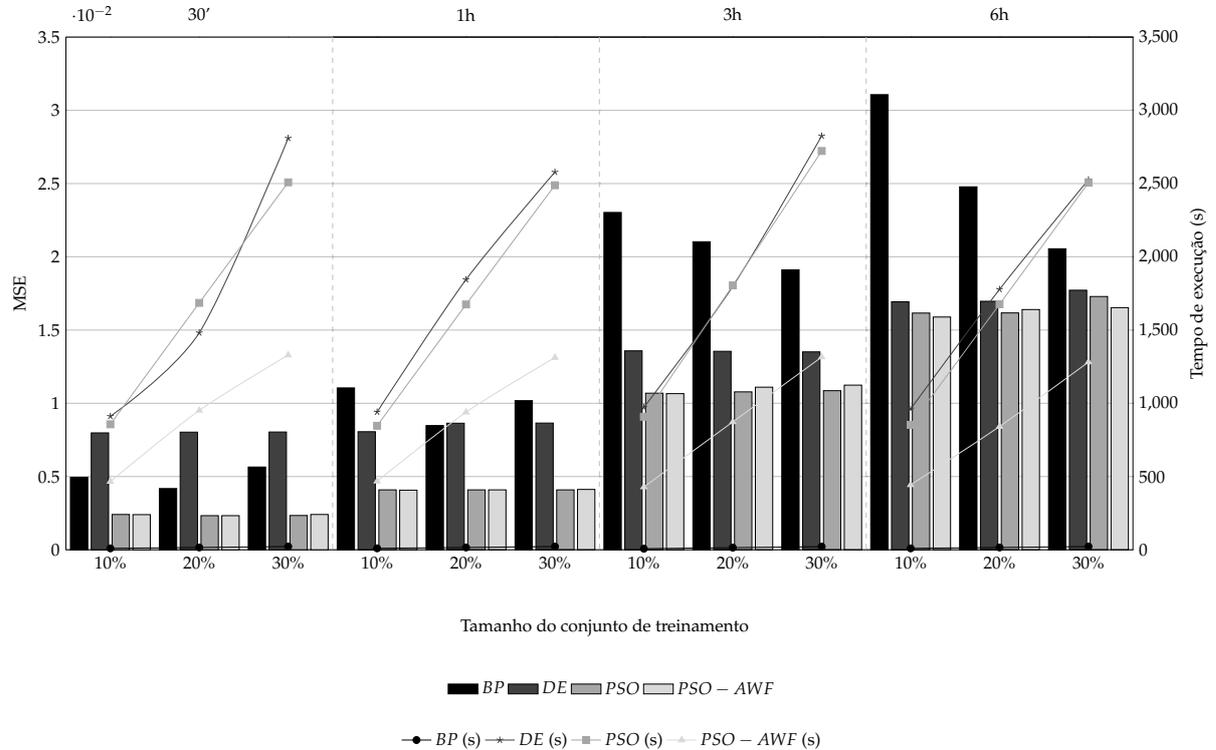
Fonte: Elaborado pelo autor (2022).

4.1.3.2 Estudo de caso 2

A Figura 13 é um diagrama que resume a relação entre o tamanho do conjunto de treinamento (10%, 20%, e 30%) com o MSE (barras) e o tempo de execução (linhas), com os quais os modelos de regressão estimam 30', 1h, 3h e 6h à frente da série temporal WS. Como esperado, o tempo de execução é diretamente proporcional ao tamanho do conjunto de treinamento. O DE, PSO, e PSO-AWF não apresentaram qualquer tipo de tendência entre o tamanho do conjunto de treinamento e o MSE. No entanto, o BP reduziu o valor do MSE com o aumento do tamanho do conjunto de treinamento para 3h e 6h, e não apresentou nenhum tipo de tendência para 30' e 1h. Esta figura mostra apenas os resultados usando a velocidade do vento como entrada da FTDNN, mas o mesmo comportamento foi encontrado com todas as

combinações possíveis com v e as outras três medições (D , p , e T).

Figura 13 – Relação entre o tamanho do conjunto de treinamento, o MSE, e o tempo de execução.



Fonte: Bezerra *et al.* (2017).

A Tabela 6 resume a relação entre o tamanho do conjunto de treinamento mais frequente, e o horizonte de previsão. Os melhores resultados calculados pelos quatro métodos de treinamento (BP, DE, PSO, e PSO-AWF) usando uma FTDNN para prever os dados do conjunto de dados WS são apresentados.

Para 30' à frente, o tamanho mais frequente do conjunto de treinamento foi de 10% para o BP, DE e PSO, e para o PSO-AWF 30%. Na sequência, para 1h à frente o tamanho mais frequente foi de 20% para o BP, DE e PSO, e 10% para o PSO-AWF. Finalmente, nos horizontes de 3h e 6h à frente o conjunto de treinamento mais frequente foi o de 30% para o BP, e 10% tanto para o PSO quanto para o PSO-AWF. Para o DE os resultados para 3h e 6h foram 30% e 10% respectivamente. Pode-se observar ainda que o BP aumentou o tamanho do conjunto de treinamento com o aumento do horizonte de previsão. Por outro lado, para o PSO e PSO-AWF os melhores resultados foram obtidos com o conjunto de treinamento menor (10%) nos horizontes de previsão de 3h e 6h, para 30' e 1h o tamanho do conjunto de treinamento não apresentou nenhuma relação entre o seu tamanho e o resultado obtido pelo modelo, sendo assim o menor conjunto escolhido por trazer o mesmo resultado que os outros.

Tabela 6 – Resultados das FTDNN com diferentes tamanhos de conjunto de treinamento.

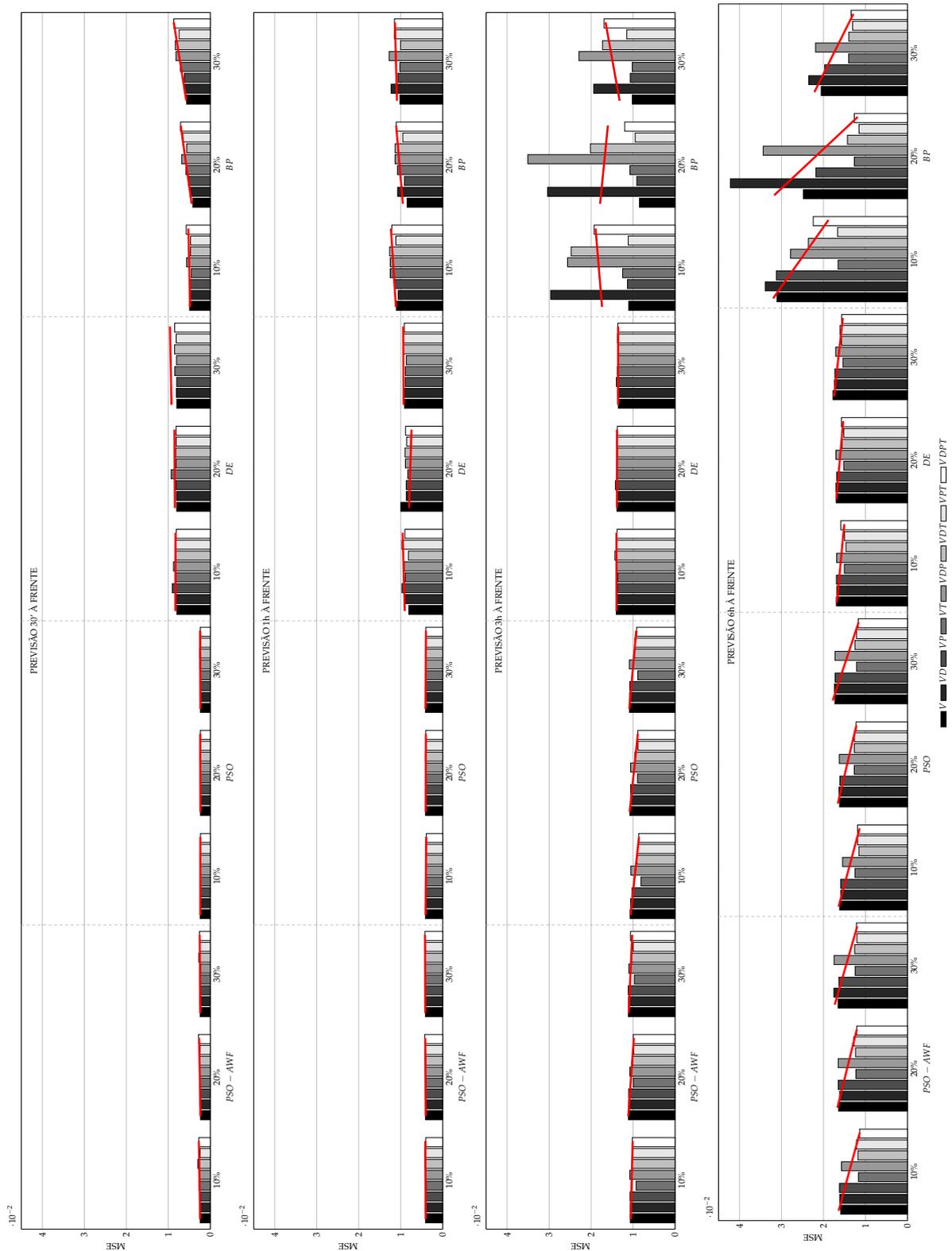
Dados	Método	30'			1h			3h			6h		
		%	t (s)	MSE (10^{-5})	%	t (s)	MSE (10^{-5})	%	t (s)	MSE (10^{-5})	%	t (s)	MSE (10^{-5})
v	BP	20	15	418	20	15	848	30	21	1911	30	22	2055
vp		10	10	485	20	16	912	30	21	2067	30	23	1972
vT		10	10	454	30	23	1018	30	21	1773	20	16	1264
vD		20	17	479	10	9	1060	30	28	1939	30	26	2351
vpT		10	10	471	20	17	950	30	22	1684	20	17	1153
vDp		10	8	561	20	15	1132	30	23	2289	30	23	2188
vDT		10	8	472	30	23	1003	30	24	1728	30	23	1395
$vDpT$		10	8	573	20	15	1106	20	15	1199	20	15	1265
v		DE	10	911	798	10	940	806	30	2825	1351	10	961
vp	20		1830	815	20	1814	866	10	976	1373	10	966	1699
vT	10		945	815	20	1809	831	10	961	1362	30	2575	1615
vD	20		1481	803	20	1844	864	20	1798	1355	20	1777	1696
vpT	30		2785	812	20	1759	858	30	2927	1371	10	955	1618
vDp	30		2810	804	30	2578	865	30	2972	1358	30	2526	1772
vDT	10		912	816	10	981	817	30	2990	1344	10	990	1635
$vDpT$	10		959	808	20	1835	883	30	2872	1363	30	2631	1767
v	PSO		20	1685	233	30	2488	409	10	908	1069	10	852
vp		20	1830	234	20	1814	401	10	920	1025	10	903	1587
vT		30	2737	231	10	908	391	10	932	814	30	2725	1213
vD		10	960	235	20	1913	403	20	1853	1053	10	953	1585
vpT		30	2915	235	20	1995	390	20	1941	881	10	965	1185
vDp		10	970	232	20	1974	405	10	1013	1052	10	974	1544
vDT		10	989	242	10	981	395	10	990	904	10	990	1157
$vDpT$		10	1049	233	10	1053	388	10	1222	863	30	3158	1173
v		PSO-AWF	20	942	237	10	468	407	10	428	1066	10	443
vp	10		474	235	10	477	405	10	475	1070	10	460	1619
vT	30		1421	238	10	478	389	10	460	929	10	459	1166
vD	20		964	242	20	974	409	10	460	1066	10	469	1614
vpT	30		1528	240	30	1544	408	20	972	982	30	1467	1202
vDp	30		1468	233	10	503	411	10	482	1077	10	485	1570
vDT	20		998	253	20	1024	406	20	980	1011	10	492	1180
$vDpT$	30		1592	256	10	526	403	20	1029	996	10	513	1139

Fonte: Adaptado de Bezerra *et al.* (2017).

O gráfico apresentado na Figura 14 sumariza a relação entre as variáveis de entrada com o MSE (barras). As linhas vermelhas são regressões simples usadas para representar as tendências do MSE. O PSO e PSO-AWF apresentaram resultados quase constantes de MSE para os horizontes de previsão de 30' e 1h, independentemente das variáveis utilizadas como entradas. Para o horizonte de 3h à frente, apresenta uma ligeira redução do MSE com o aumento do número de variáveis utilizadas como entradas. Finalmente, para 6h o aumento no número de variáveis de entrada leva à redução do MSE. No que diz respeito ao BP, com o aumento do número de variáveis de entrada o MSE também aumentou para os horizontes de 30' e 1h. Para o horizonte de 3h à frente, não se observa nenhuma tendência consistente. Finalmente, para 6h o aumento no número de variáveis de entrada leva à redução do MSE. Em geral, o aumento do horizonte de previsão requer mais informação.

A Figura 15 é usada para analisar a influência do NNHL (3, 5, 7, 11) e LAG (1, 3, 5, 7) no MSE para o horizonte de tempo de 30', para cada método de aprendizado (BP, PSO e PSO-AWF), com a variável de entrada v , e o tamanho conjunto de treinamento de 10%. O melhor resultado de cada combinação NNHL e LAG é apresentado. Cada rodada considera uma

Figura 14 – Relação entre o número de variáveis de entrada e o MSE.

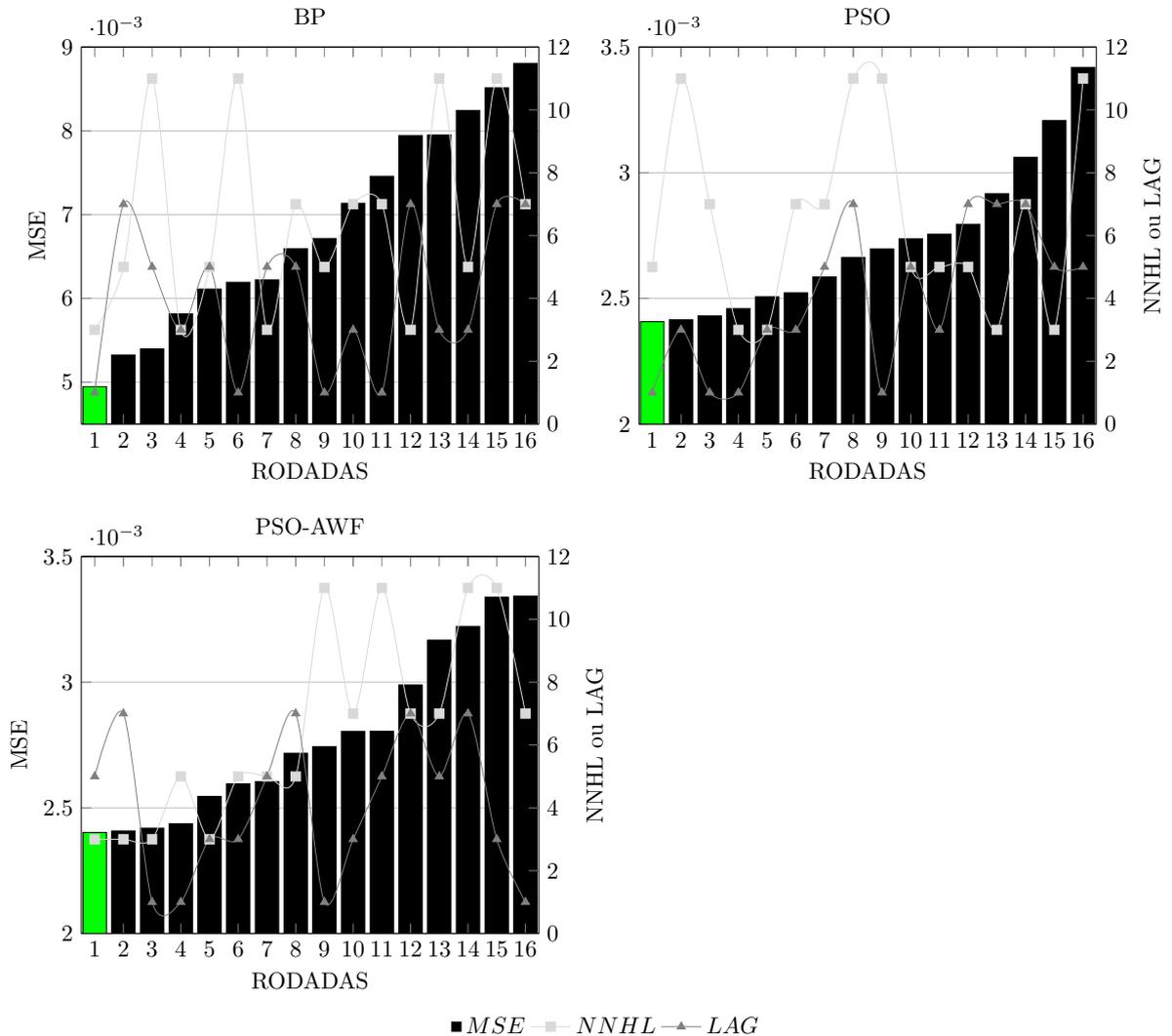


Fonte: Adaptado de Bezerra *et al.* (2017).

das combinações possíveis entre NNHL e LAG. O PSO-AWF apresentou o mesmo valor de NNHL para os 3 melhores resultados e não foi possível identificar quaisquer tendências para o

BP e PSO.

Figura 15 – Relação entre o NNHL, LAG, e MSE.



Fonte: Elaborado pelo autor (2021).

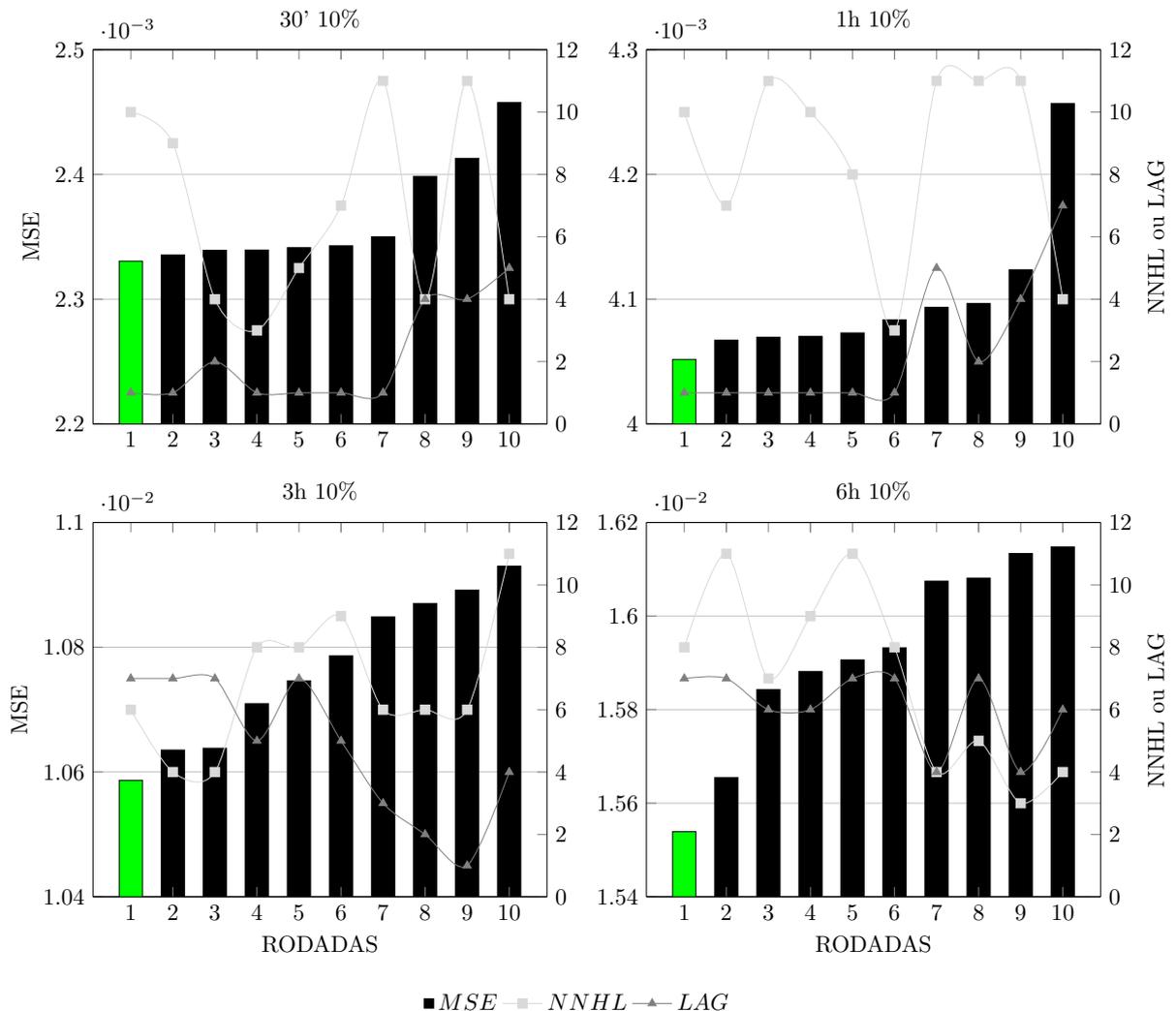
4.1.3.3 Estudo de caso 3

Considerando os resultados obtidos por TE na Seção 4.1.3.2, considera-se v como a única variável de entrada para os horizontes de previsão de 30', 1h e 3h, e combinações de v com as outras variáveis são empregadas como entradas para 6h à frente, o tamanho do conjunto de treinamento será sempre 10%.

A Figura 16 é usada para analisar a influência do $NNHL \in \mathbb{Z} | 3 \leq NNHL \leq 11$, e $LAG \in \mathbb{Z} | 1 \leq LAG \leq 7$ no MSE para o PSO-AWF como método de aprendizagem e escolha desses parâmetros da FTDNN. Os resultados de 10 rodadas são apresentados. A partir da Figura 15 não foi possível verificar uma tendência entre o NNHL e LAG com o MSE. O PSO-AWF mostra que,

para previsões de 30' e 1h, os valores de LAG tendem a ser pequenos (geralmente um), o oposto foi encontrado para os horizontes de 3h e 6h onde o valor mais frequente é 7. Os valores de NNHL para horizontes de 1h e 6h foram altos e para 30' e 3h, não foi possível observar qualquer tendência.

Figura 16 – Relação entre o NNHL, LAG, e MSE.

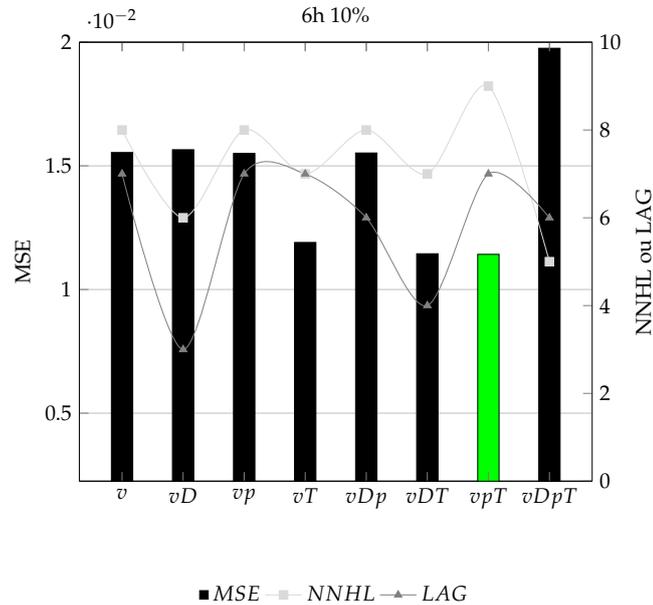


Fonte: Elaborado pelo autor (2021).

Para o horizonte de 6h, a Figura 16 apresenta os resultados dos modelos que usam apenas v como entrada. A Figura 17 apresenta os melhores resultados para cada conjunto de dados utilizados como entrada. Os melhores resultados foram obtidos utilizando v , p e T . O comportamento dos resultados foi similar aos encontrados através da TE na Figura 14. Independente das variáveis de entrada utilizadas, o valor do NNHL foi alto (normalmente 8), e o valor mais frequente para o LAG foi 7.

A Figura 18 foi usada para verificar se o uso do PSO-AWF para buscar os valores do

Figura 17 – Melhores resultados para cada conjunto de variáveis de entrada.



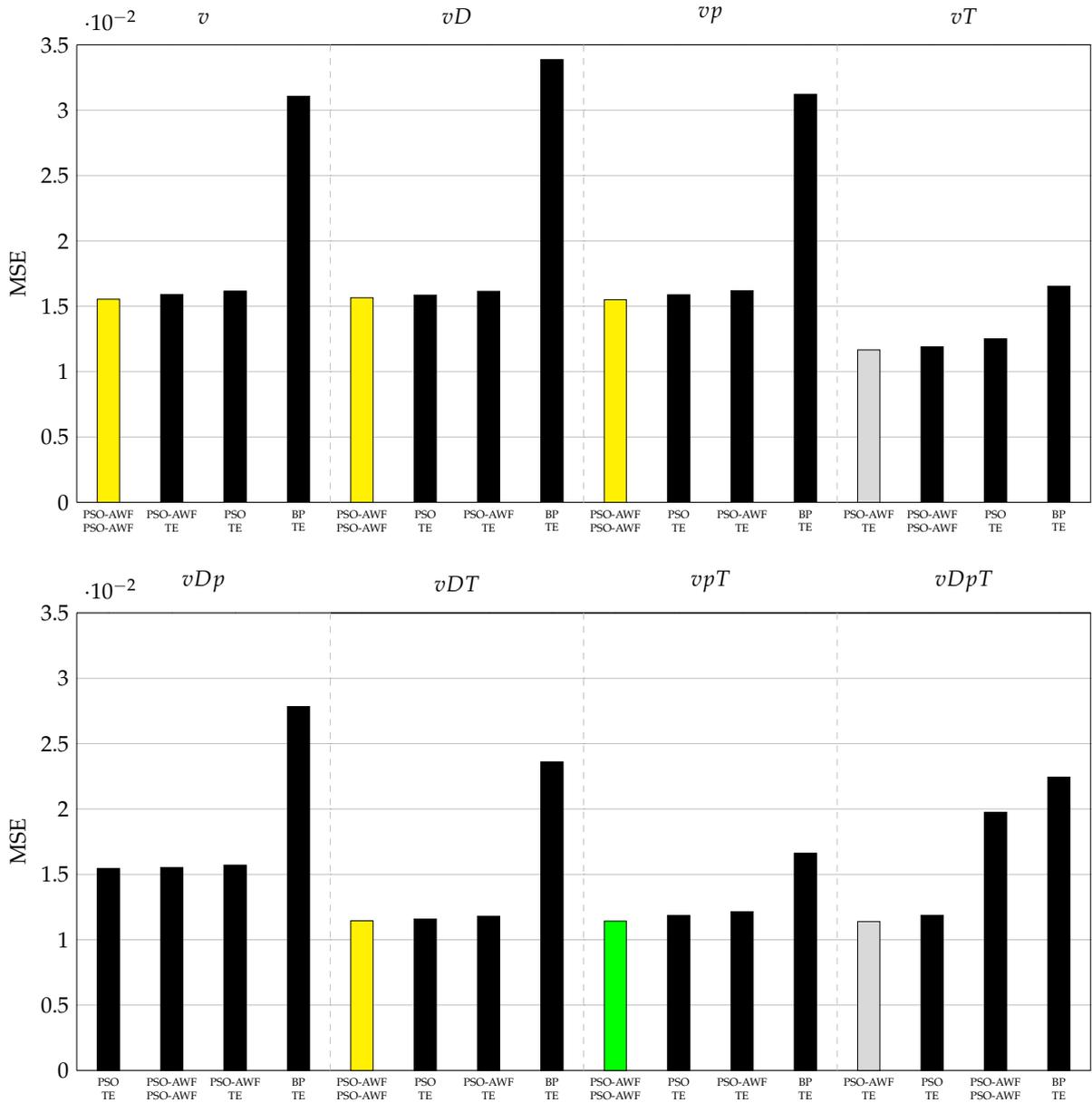
Fonte: Elaborado pelo autor (2021).

NNHL e LAG produziria melhores resultados que os encontrados através da TE, para cada tipo de variável de entrada disponível (v, D, p, T), bem como suas combinações utilizando cada algoritmo de treinamento. O PSO-AWF obteve o melhor resultado (barra verde), e esteve presente em 7 das 8 melhores soluções. Cinco destas soluções usaram o PSO-AWF como método de treinamento e escolha do valor do NNHL e LAG (barras amarelas), para as outras duas (barras cinzas) o PSO-AWF foi usado como método de treinamento e a escolha do valor do NNHL e LAG feita através da TE.

Para todos os horizontes de previsão, a Figura 19 foi usada para verificar se o uso do PSO-AWF para buscar os valores do NNHL e LAG produziria melhores resultados que os encontrados através da TE, para cada tipo de variável de entrada disponível. Mesmo com a ampliação da busca: NNHL de 3, 5, 7, e 11 para $NNHL \in \mathbb{Z} | 3 \leq NNHL \leq 11$, e LAG de 1,3,5, e 7 para $LAG \in \mathbb{Z} | 1 \leq LAG \leq 7$; a restrição no tamanho do conjunto de treinamento (10%); a redução do número de rodadas de 32 para 10; e para os horizontes de 30', 1h, e 3h, a restrição na quantidade de variáveis utilizadas como entrada (apenas v), o PSO-AWF obteve o melhor resultado (barra verde) entre todos os horizontes.

Para um melhor entendimento, o tempo de execução não foi apresentado na última figura porque não é mais tratado como uma média, e sim como o tempo acumulado pelo experimentos 2 e 3. No Estudo de Caso 2 onde apenas o v foi considerado como entrada, e o tamanho do conjunto de treinamento 10%, obteve-se para o BP, PSO, e PSO-AWF aproximadamente 0.16,

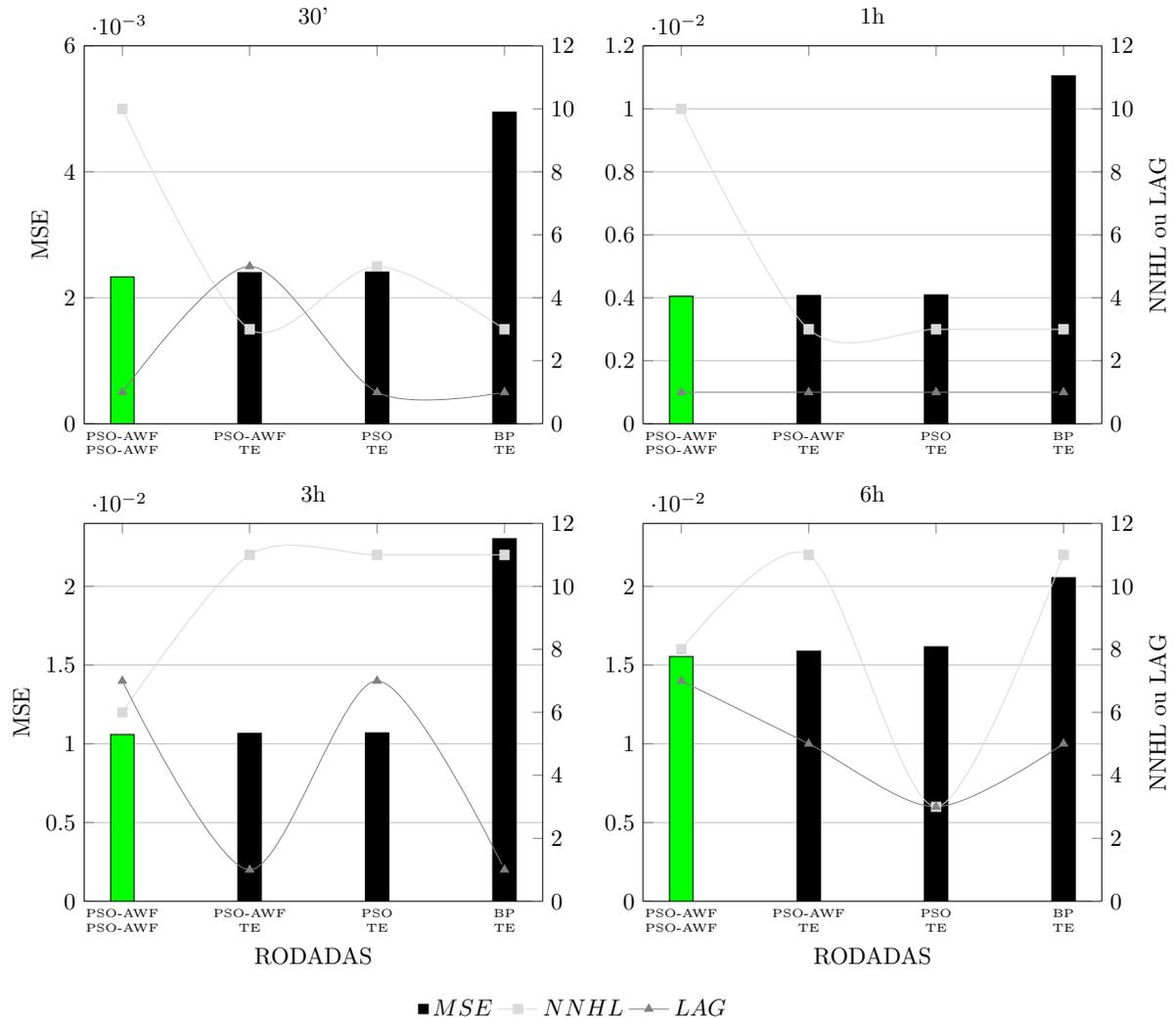
Figura 18 – Performance da FTDNN para diferentes dados de entrada e algoritmos de treinamento usados para obter os valores de NNHL e LAG.



Fonte: Elaborado pelo autor (2021).

7.93, e 3.96 horas para a TE respectivamente. Extrapolando este cenário para as mesmas condições usadas pelo PSO-AWF, com NNHL = [3,11] e LAG = [1,7], obteve-se aproximadamente 0.64, 31.24, e 15.6 horas. No Estudo de Caso 3, o PSO-AWF gastou cerca de 12.85 horas para executar a mesma tarefa do Estudo de Caso 2. Uma redução de quase 58.9% quando comparado ao PSO, e 17.6% ao PSO-AWF ambos usando o TE para encontrar NNHL e LAG.

Figura 19 – Performance do PSO-AWF comparado aos métodos de aprendizado usando TE.



Fonte: Elaborado pelo autor (2021).

4.2 Método online

Nesta seção, o comportamento de busca da máquina *kernel* com múltiplos desafios resolvendo problemas de regressão é estudado através de experimentos. O MC-KRLS é aplicado ao banco de dados de geração eólica, estimando $t + 5$ minutos à frente. Os efeitos da abordagem auto adaptativa serão analisados. O desenvolvimento que se segue objetiva responder duas importantes questões: Como o MC-KRLS funciona em problemas de regressão, e como os parâmetros do MC-KRLS são melhores ajustados.

Todos os algoritmos foram escritos em Inc. (2014). O teste e validação das técnicas propostas foram realizados no computador de alta performance, do inglês *high performance computer* (HPC), da universidade técnica da Dinamarca, do dinamarquês *Danmarks Tekniske Universitet* (DTU), *clusters* central e compute, parceria viabilizada através do programa Ci-

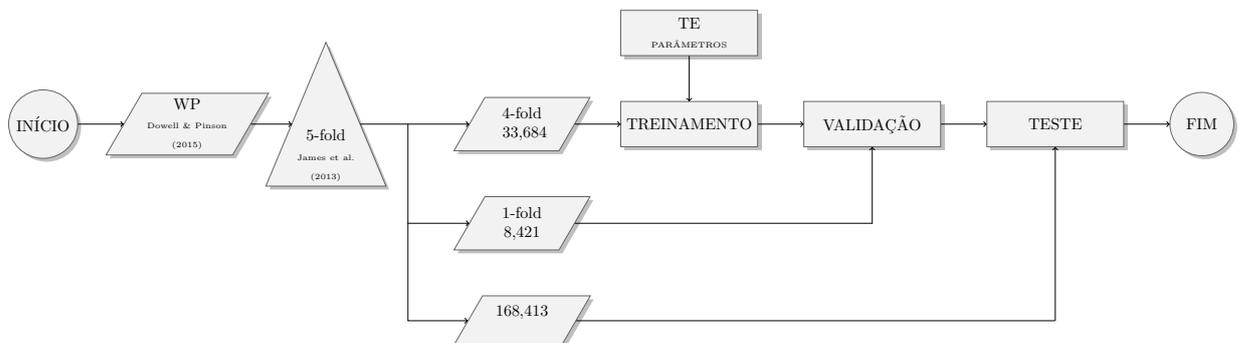
ência sem Fronteiras, Cronograma de Projeto 204899/2014-1, sob contrato de concessão no. 8499660428694715.

Para realizar o estudo comparativo e verificar a performance do método proposto, cinco modelos são usados como referência. Além do modelo de Persistência já apresentado na Seção 2.3.1, e as três máquinas *kernel* apresentadas na Seção 3.3.1, o *online sequential extreme learning machine* (OS-ELM) é utilizado para comparação. Este modelo foi escolhido por ser extremamente rápido e os dados utilizados no treinamento serem apresentados sequencialmente para o algoritmo de aprendizado. A aplicação sequencial da solução dos mínimos quadrados do vetor de pesos de saída (β) resulta no OS-ELM que usa o algoritmo RLS (CHONG; ZAK, 2013). O OS-ELM e o ELM (HUANG *et al.*, 2011) são essencialmente o mesmo algoritmo, e podem atingir a mesma performance de aprendizado quando o número de dados do modo *batch* é o mesmo do conjunto de dados de treinamento.

A Figura 20 apresenta diagramas de blocos dos processos dos estudos de caso 4 e 5. Apontando o conjunto de dados utilizado (Seção 4.2.1), como os dados foram extraídos e divididos (Seção 4.2.2), as ferramentas utilizadas no processo de treinamento (Seção 3.3), e por fim o porque dos diferente arranjos (Seção 4.2.3). Note que as etapas dos dois estudos de caso são as mesmas, a diferença são os modelos aplicados

- estudo de caso 4, *test case 4* (TC4), persistência e oito estruturas compostas pelos *single input and single output* (SISO) e *multiple input and multiple output* (MIMO) dos modelos OS-ELM, ALD-KRLS, SW-KRLS e KRLS-T; e
- estudo de caso 5, *test case 5* (TC5), vinte e quatro estruturas compostas pelos M2C e M3C dos modelos ALD-KRLS, SW-KRLS e KRLS-T, aplicando o método de aprendizado descrito na Seção 3.4.

Figura 20 – Diagrama de blocos dos processos dos estudos de caso 4 e 5.



Fonte: Elaborado pelo autor (2021).

4.2.1 Conjunto de dados

Os dados foram utilizados por Dowell e Pinson (2016) e são compostos por medições da potência eólica gerada a cada 5 minutos, de 01/01/2011 04h05 até 01/01/2013 04h00, de 23 parques eólicos *onshore*, resultando em 210.528 medições para cada parque disponibilizado pelo AEMO (2021). O conjunto de dados completos está disponível para *download* em Dowell (2015).

Todas as plantas possuem capacidade máxima entre 30 MW e 420 MW. A Tabela 7 apresenta uma análise do conjunto de dados; ‘Min’, ‘Méd’, e ‘Max’ são, respectivamente, os valores mínimo, médio, e máximo encontrados, ‘NaN’ é o número de medições não disponíveis, ‘Zeros’ é o número de dados com valor igual a zero, e ‘Neg’ é o número de dados com valores negativos. Os valores negativos representam o consumo das turbinas quando não estão gerando potência.

Tabela 7 – Informações sobre cada parque eólico.

Parque Eólico	Min	Méd	Max	NaN	Zeros	Neg
MTMILLAR	-1.01	21.60	69.46	1	1376	5659
WATERLWF	-1.20	38.02	111.0	0	735	12878
WOOLNTH1	-1.53	56.73	138.5	1	2250	11505
WAUBRAWF	-1.90	78.91	185.8	1	3771	12658
HALLWF2	-0.82	32.19	70.94	0	975	18826
SNOWTWN1	-1.07	46.77	99.10	0	1912	18270
HALLWF1	-1.17	41.81	93.27	0	1949	18412
LKBONNY2	-2.25	46.83	156.9	0	89	21084
NBHWF1	-1.60	57.53	131.8	0	1381	20155
CULLRGWF	-9.98	12.34	33.67	1	902	21507
STARHLWF	-19.79	9.98	34.49	1	6404	16953
YAMBUKWF	-1.22	11.06	34.00	1	8098	15572
LKBONNY1	-1.43	24.57	79.40	1	1453	22932
WPWF	-1.00	35.04	92.00	1	28408	9
CLEMGPF	-3.20	22.68	137.0	0	2625	27273
CATHROCK	0.00	23.62	62.23	1	30708	0
LKBONNY3	-0.67	11.72	38.52	0	6076	32496
GUNNING1	-10.00	16.89	50.00	23856	4242	18146
BLUFF1	-0.30	13.00	53.30	53232	10575	1958
WOODLWN1	-3.62	14.67	47.90	35088	38233	162
CAPTLWF	-10.00	54.55	138.0	1	72889	919
OAKLAND1	-9.62	15.14	66.10	62160	40847	654
MACARTH1	-3.90	16.63	391.6	134112	23628	946

Fonte: Elaborado pelo autor (2021).

Para esta pesquisa, oito parques eólicos (nomes em negrito) foram escolhidas por terem o menor número de medições não disponíveis (0). Todos os valores negativos foram considerados iguais a zero, e os dados foram normalizados entre]0,1[.

4.2.2 Definições de parâmetros

Os melhores parâmetros foram encontrados através do método de validação cruzada *k-fold* com 5 blocos de 8421 medições. Um desses blocos foi deixado de fora do processo de aprendizado e utilizado como conjunto de validação, e o restante dos dados (168413 medições) foram usados como conjunto de teste (JAMES *et al.*, 2013). Para ser justo, as mesmas partições dos dados de treinamento, validação e teste são usadas quando diferentes algoritmos são treinados no conjunto de dados. Observa-se que não existe uma regra de divisão de dados, apenas práticas comuns. A raiz do erro médio quadrático percentual, *root mean square percentage error* (RMSPE), é usado como o índice de desempenho.

O parâmetro que controla o tamanho do dicionário para cada algoritmo determina as complexidades computacionais e de memória. Para cada experimento, este parâmetro é variado em uma ampla faixa para que o RMSE possa ser medido. Curvas de desempenho são usadas para fins de comparação. A Seção 4.2.3.2 apresenta as curvas do *benchmark* fornecidas pelo ALD-KRLS e seus respectivos MC-KRLS como parte da análise dos parâmetros *approximate linear dependency* (ALD) *threshold* (ν) e o comprimento ótimo do LAG.

Note que a seleção ótima dos parâmetros é um problema importante em aberto na literatura de máquinas *kernel*, mas está fora do escopo desta pesquisa. Os parâmetros restantes são escolhidos através de uma busca exaustiva com o objetivo de otimizar a curva de performance, e são listados na Tabela 8. Onde M é o tamanho do dicionário, σ é o parâmetro de largura do *kernel*, c é um parâmetro de regularização, λ é o fator de esquecimento, *jitter* é o ruído utilizado para evitar o erro de arredondamento, e $sn2$ é uma relação entre o ruído e o sinal. Note que o NNHL é uma adaptação do termo *number of hidden nodes* (NOHN).

4.2.3 Resultados

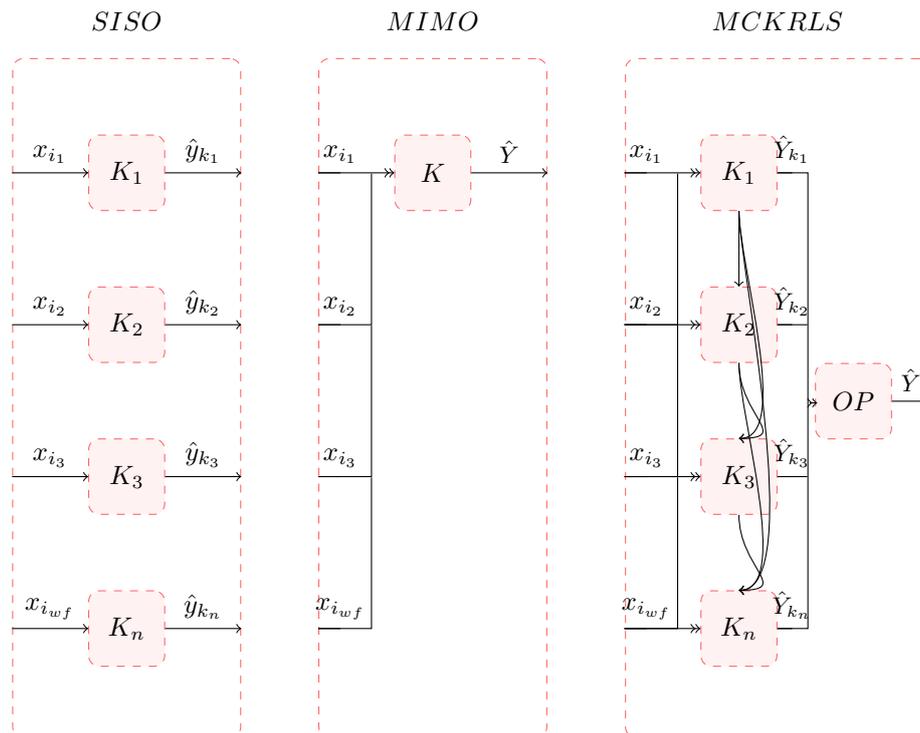
A Figura 21 ilustra os modelos utilizados nos dois estudos de caso, nomeados TC4, composto pelos blocos tracejados SISO e MIMO, e TC5, representado pelo MC-KRLS, onde o algoritmo proposto com n máquinas *kernel* MIMO rastreando os resultados de cada *kernel*, apresenta o resultado para um ou vários parques eólicos. A caixa ‘OP’ representa o ‘operador’ que irá combinar todas as previsões.

Tabela 8 – Parâmetros usados pelos modelos.

Modelo	Parâmetro
ALD-KRLS	$M = 20, 30, 40, 50, 60, 70$
SW-KRLS	$LAG = 1, 2, 3, 4, 5, 6, 7, 8$
KRLS-T	<i>Gaussian Kernel</i> Parâmetro do <i>kernel</i> (σ) = 32
ALD-KRLS	$v = 2E-5, 1E-5, 2E-4, 1E-4, 1E-3, 0.05, 0.01, 0.1$
SW-KRLS	$c = 2E-5, 1E-5, 2E-4, 1E-4, 1E-3, 0.05, 0.01, 0.1$
KRLS-T	$\lambda = 0.9, 0.99, 0.999$ $jitter = 5E-6, 1E-6, 1E-5$ $sn2 = 5E-3, 5E-2, 1E-2$
OS-ELM	$NNHL = 3, 5, 10, 20, 30, 50, 75, 100, 150, 200, 250$ $LAG = 1, 2, 3, 4, 5, 6, 7, 8$ <i>Radial Basis Function</i> (RBF) Número de dados de treinamento = 33,684

Fonte: Bezerra *et al.* (2021).

Figura 21 – Diagrama de blocos com todos os arranjos das máquinas KRLS aplicadas nos experimentos.



Fonte: Bezerra *et al.* (2021).

4.2.3.1 Estudo de caso 4

A Tabela 9 apresenta o RMSPE de cada parque eólico. A última linha contém suas respectivas médias. As diferenças entre máquinas *kernel* SISO são pequenas. No entanto, a máquina *kernel* que usa a ALD como ferramenta de esparsidade é a melhor.

O uso de arranjos MIMO melhora o RMSPE quando comparado com suas contrapartes SISO, exceto para o SW-KRLS. O MIMO ALD-KRLS prova ser a melhor máquina *kernel* novamente. O uso de MIMO para o SW-KRLS apresenta RMSPE ligeiramente superior quando comparado com as respectivas versões SISO. O KRLS-T apresenta resultados iguais devido ao erro de arredondamento, mas a versão MIMO apresenta melhor desempenho em sete dos oito sites. A exceção é o parque eólico LKBONNY3, que faz com que a média seja ligeiramente superior à da versão SISO.

Tabela 9 – RMSPE dos modelos de referência.

Parque Eólico	Persistência	OS-ELM		ALD-KRLS		SW-KRLS		KRLS-T	
		SISO	MIMO	SISO	MIMO	SISO	MIMO	SISO	MIMO
WATERLWF	3.863	3.854	3.879	3.853	3.833	3.959	4.223	3.860	3.859
HALLWF2	3.846	3.837	3.821	3.836	3.767	3.999	4.171	3.852	3.851
SNOWTWN1	3.393	3.352	3.411	3.357	3.341	3.566	3.804	3.409	3.408
HALLWF1	3.450	3.431	3.388	3.433	3.342	3.577	3.727	3.461	3.461
LKBONNY2	3.760	3.678	3.712	3.679	3.614	3.817	4.077	3.777	3.776
NBHWF1	3.000	2.916	2.937	2.926	2.877	3.117	3.273	3.029	3.028
CLEMGPWF	1.893	1.884	1.915	1.888	1.876	2.064	2.220	1.895	1.894
LKBONNY3	5.639	5.552	5.573	5.545	5.514	5.745	6.039	5.618	5.622
MÉDIA	3.606	3.563	3.580	3.565	3.520	3.731	3.942	3.613	3.613

Fonte: Bezerra *et al.* (2021).

O coeficiente de correlação de Pearson (R) é usado para mostrar porque alguns modelos MIMO alcançam resultados melhores do que aqueles obtidos com modelos SISO. Se R for maior que 0,8, então o sistema é descrito como fortemente correlacionado, ao passo que quando R é menor que 0,5 é fracamente correlacionado. A Tabela 10 apresenta todos os resultados. As últimas três linhas apresentam um resumo dos parâmetros classificados como fortemente, fracamente, e correlacionados. Cinco dos oito parques estão fortemente ou correlacionados com quatro ou mais parques eólicos.

Tabela 10 – Coeficiente de correlação de pearson

Parque Eólico	WATERLWF	HALLWF2	SNOWTWN1	HALLWF1	LKBONNY2	NBHWF1	CLEMGPWF	LKBONNY3
WATERLWF	1							
HALLWF2	0.82	1						
SNOWTWN1	0.41	0.53	1					
HALLWF1	0.79	0.91	0.49	1				
LKBONNY2	0.29	0.29	0.10	0.33	1			
NBHWF1	0.79	0.87	0.54	0.90	0.26	1		
CLEMGPWF	0.50	0.54	0.77	0.50	0.05	0.63	1	
LKBONNY3	0.36	0.34	0.14	0.37	0.90	0.29	0.08	1
FORTEMENTE C.	1	3	0	2	1	2	0	1
CORRELACIONADO	3	2	3	2	0	3	5	0
FRACAMENTE C.	3	2	4	3	6	2	2	6

Fonte: Bezerra *et al.* (2021).

Considerando todos os modelos, o uso de máquinas MIMO gerou resultados similares quando comparados às suas respectivas máquinas SISO, mas o ALD-KRLS apresentou uma melhora no RMSPE e foi a melhor máquina *kernel* novamente. Para o horizonte de 5 minutos, o uso de máquinas SW-KRLS e KRLS-T MIMO apresentaram o RMSPE ligeiramente mais alto comparado às suas versões SISO. O KRLS-T mostra resultados iguais por causa de erros de arredondamento.

Os melhores parâmetros foram encontrados através do procedimento de validação cruzada *k-fold* usando os valores apresentados na Tabela 8. Em todos os casos que v e c são maiores que 0.001 existiu um aumento significativo do RMSPE. Além disso, em todos os casos, quanto menor o tamanho do dicionário e maior o LAG, maior o RMSPE. Finalmente, quando usando o ALD-KRLS, duas séries temporais foram fortemente afetadas pelo tamanho do dicionário: HALLWF1 e NBHWF1. Os autores do KRLS-T afirmam que usualmente, λ é sensível entre $[0.95, 1]$ (VAERENBERGH *et al.*, 2012), e nesta pesquisa os melhores valores de RMSPE foram alcançados usando λ igual a 0.9 e o erro cresce ao aumentar seu valor.

4.2.3.2 Estudo de caso 5

As diferenças entre os modelos de referência, SISO e MIMO, e os modelos MC-KRLS são pequenas. Porém, os resultados obtidos com os modelos MC ALD sugeridos são, em geral, melhores do que os do SISO e MIMO ALD-KRLS, conforme a Tabela 11. Os modelos SW-KRLS e KRLS-T agora apresentam o mesmo comportamento anteriormente observado nos modelos SISO e MIMO (TC4).

Tabela 11 – RMSPE de todos os modelos

	SISO		MIMO							
	KRLS	KRLS	M2C KRLS	M2C KRLSW	M2C KRLSA	M2C KRLSAW	M3C KRLS	M3C KRLSW	M3C KRLSA	M3C KRLSAW
ALD-KRLS	3.565	3.520	3.522	3.520	3.522	3.520	3.519	3.517	3.520	3.519
SW-KRLS	3.731	3.942	3.953	3.920	3.948	3.928	3.921	3.925	3.909	3.914
KRLS-T	3.613	3.613	3.678	3.652	3.661	3.641	3.649	3.636	3.645	3.636

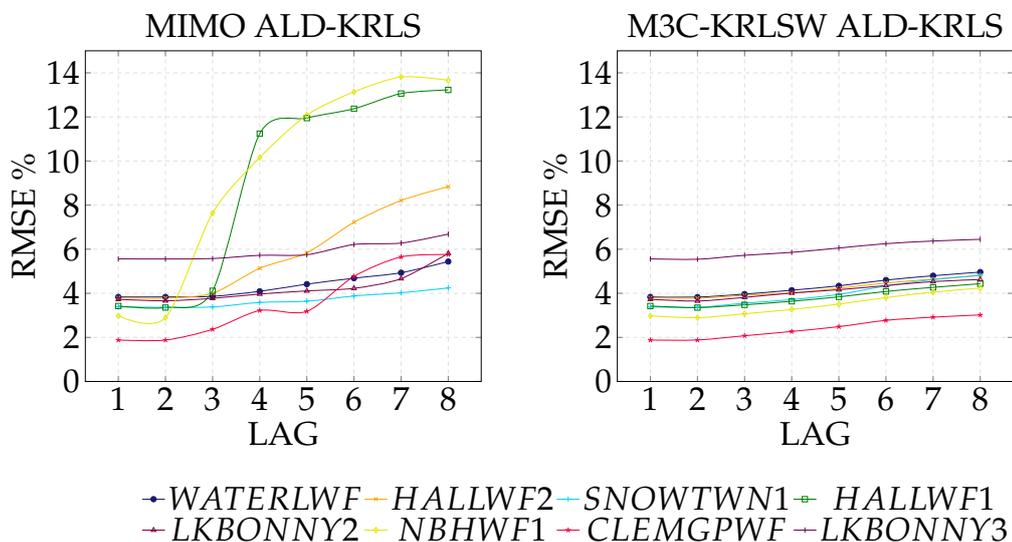
Fonte: Bezerra *et al.* (2021).

Analogamente ao observado com as máquinas ALD-KRLS SISO e MIMO (TC4), há um aumento significativo do RMSPE em todos os casos em que v é maior que 0,001, mas se verificam outros comportamentos diferentes. Em primeiro lugar, o tamanho do dicionário ou o número de LAG usados não influenciam significativamente o RMSPE, permitindo assim o uso do menor tamanho de dicionário e do menor comprimento de LAG. Em segundo lugar, ao usar o

ALD-KRLS, duas séries temporais são significativamente afetadas pelo tamanho do dicionário, HALLWF1 e NBHWF. No entanto, isso não é observado no caso de KRLS múltiplos. Terceiro, o uso de qualquer um dos modelos de múltiplos desafiadores mostra resultados mais estáveis em comparação com o ALD-KRLS.

A Figura 22 mostra os resultados de MIMO e M3C-KRLSW ALD-KRLS, que são os melhores resultados apresentados nas Tabelas 9 e 11. Observa-se que o uso de múltiplos *kernels* minimiza a importância do número de LAG para qualquer valor v .

Figura 22 – Comparação entre o ALD-KRLS MIMO e ALD-KRLS M3C-KRLSW, $v = 1.00E-05$ e variando o número de LAG.



Fonte: Bezerra *et al.* (2021).

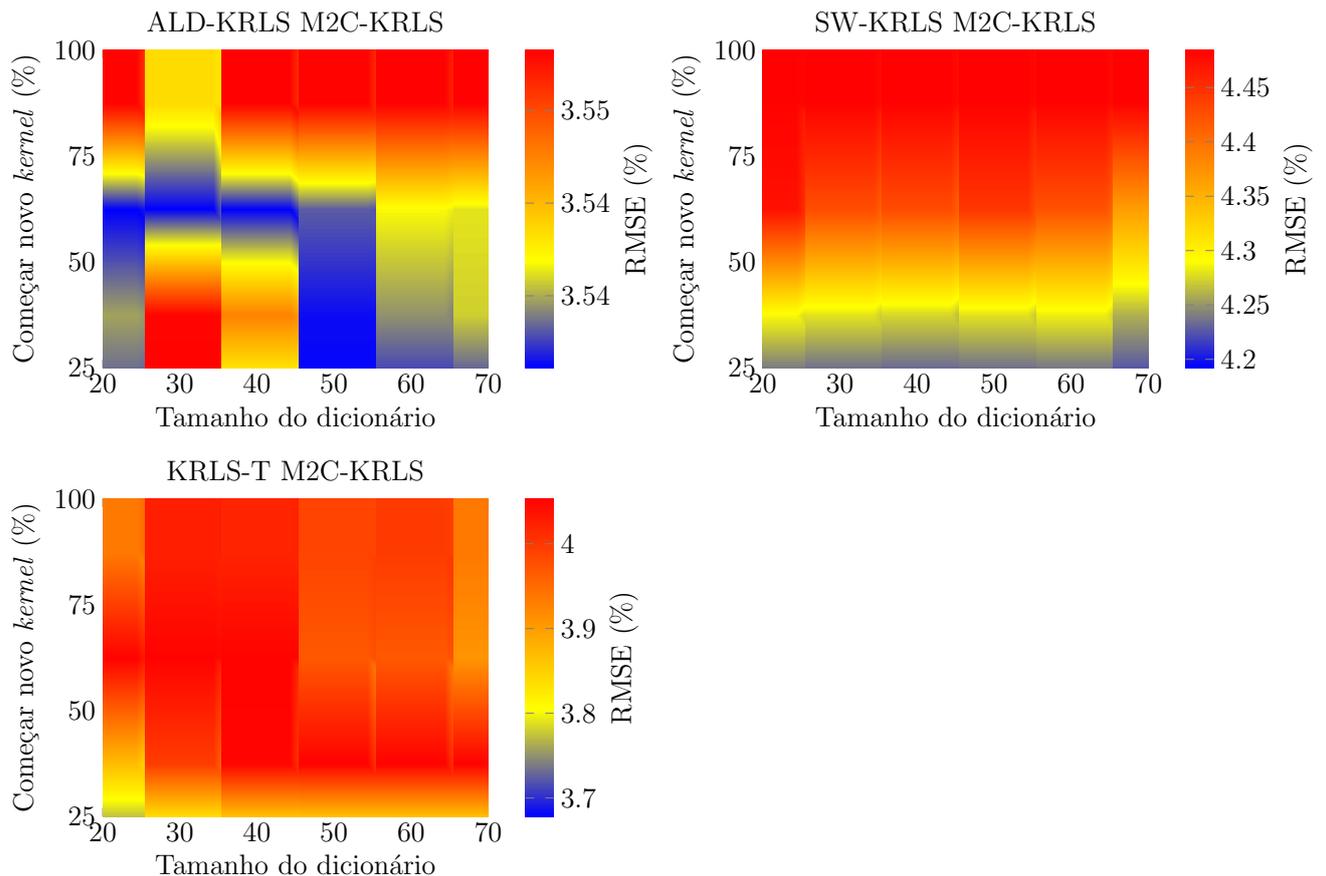
O SW-KRLS MC-KRLS apresenta um aumento significativo no RMSPE em todos os casos em que c é maior que 0,001. Para todas as condições, quanto menor o tamanho do dicionário e maior o LAG, maior o RMSPE. Para o KRLS-T, o melhor RMSPE geralmente é alcançado usando $\lambda = 0,9$, enquanto o erro aumenta com o aumento de λ .

A primeira questão levantada na Seção 3.4.1 pode ser respondida com a ajuda da Figura 23, que é usada para analisar o comportamento do RMSPE e a necessidade de eventualmente iniciar uma nova máquina *kernel*. Os melhores resultados das máquinas com múltiplos desafiadores usando diferentes tamanhos de dicionário e gatilhos para iniciar uma nova máquina *kernel* são apresentados em termos de escalas de cores distintas para cada gráfico. É um gráfico de matriz que produz uma rede preenchida de retângulos sombreados, onde cada posição da matriz corresponde a um retângulo. O eixo 'Tamanho do dicionário' representa o número máximo de exemplos (20, 30, 40, 50, 60, 70) salvos em cada dicionário do *kernel*. O

eixo 'Iniciar novo *kernel*' mostra a porcentagem do número máximo de exemplos alcançados por um dicionário do *kernel* (25%, 50%, 75%, 100%) antes de um novo dicionário ser criado. Finalmente, a coluna RMSPE relaciona a escala de cores com o valor do erro.

Os melhores resultados para ALD-KRLS M2C-KRLS são o gatilho de 75% do tamanho do dicionário para criar uma nova máquina com dicionários de 20, 30 ou 40 exemplos, e um gatilho de 25% para 50, 60 e 70. O mesmo comportamento é observado em todos os demais modelos apresentados neste trabalho, conforme denotado na Tabela 3. As máquinas com múltiplos desafios que usam SW-KRLS e KRLS-T sempre criam uma nova máquina de *kernel* quando o dicionário anterior atinge 25% de seu tamanho, independente do tamanho do dicionário.

Figura 23 – Comparação entre o RMSPE versus quando iniciar uma nova máquina *kernel*, variando o tamanho máximo do dicionário.



Fonte: Bezerra *et al.* (2021).

Para responder à segunda questão levantada na Seção 3.4.2, duas soluções são apresentadas como método de cálculo da estimativa combinada (\hat{Y}): a média aritmética e a ponderada. A versão ponderada é a melhor em 10 dos 12 modelos propostos, conforme

observado na Tabela 11. As únicas exceções são o SW-KRLS M3C-KRLS comparado com o M3C-KRLSW e o SW-KRLS M3C-KRLSAW com o M3C-KRLSA.

Para a terceira questão da Seção 3.4.3, duas opções também são apresentadas para definir quando excluir uma máquina *kernel*: uma usando o MSE e outra o AE. A Tabela 11 mostra que não há diferença entre os resultados do ALD-KRLS usando duas máquinas *kernel*. Quando três máquinas *kernel* são adotadas, o uso do MSE fornece melhores resultados. Para o SW-KRLS utilizando a média aritmética, o AE também apresenta melhor desempenho. Por fim, o KRLS-T apresenta melhores resultados utilizando o MSE em todos os modelos propostos.

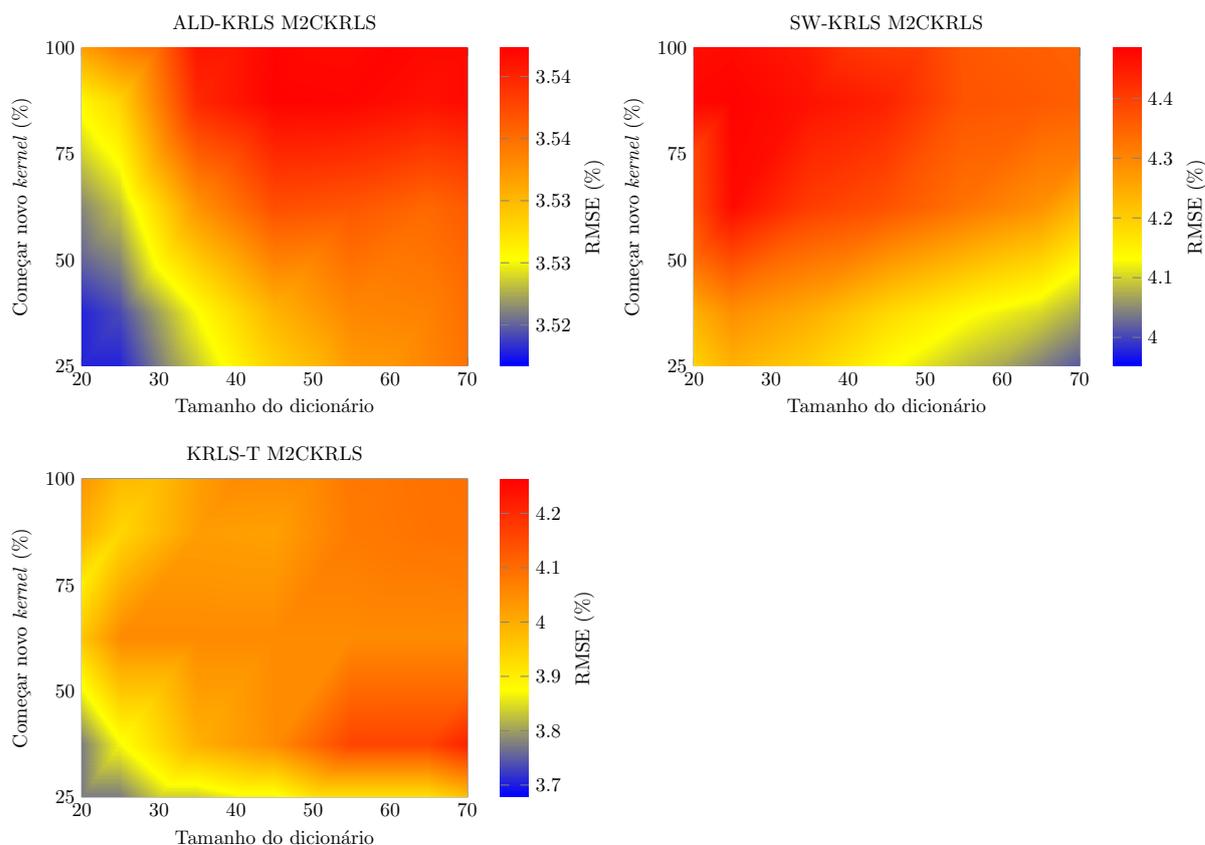
A Figura 24 apresenta o comportamento do RMSPE *versus* o tamanho do dicionário usado pelas máquinas *kernel*. O melhor resultado de várias máquinas usando diferentes tamanhos de dicionário e gatilhos para iniciar uma nova máquina *kernel* é exibido. As escalas de cores são diferentes para cada gráfico. A Figura 24 tem a mesma descrição da Figura 23. Os melhores resultados para o ALD-KRLS M2C-KRLS são encontrados com dicionário menores. As múltiplas máquinas *kernel* que utilizam o KRLS-T apresentam comportamento semelhante aos obtidos com o ALD-KRLS. As máquinas que usam o SW-KRLS obtêm os melhores resultados dicionários maiores.

Na Figura 23 e 24 é observado que o ALD-KRLS M2C-KRLS obtém o menor RMSPE, normalmente quando valores entre 25% e 75% do tamanho do dicionário é definido como gatilho para iniciar um novo *kernel*, e usa dicionários o menor dicionário (20 exemplos), a menor faixa analisada neste estudo (20, 30, 40, 50, 60, 70). Para máquinas SW-KRLS M2C-KRLS, os melhores resultados são geralmente com 25% como gatilho e o maior tamanho de dicionário (70). Para o KRLS-T M2C-KRLS, temos gatilho de 25% e o menor tamanho do dicionário. Comportamento semelhante é encontrado ao usar as máquinas listada na Tabela 3.

Para comparar o tempo computacional era necessário executar todos os métodos em uma única máquina. Os resultados apresentados antes foram de um HPC composto por *hardwares* diferentes. Em seguida, foi definido: (i) O objetivo de rastrear e analisar a métrica do *software* é determinar o tempo computacional de cada método; (ii) Cada *benchmark* era executado no mesmo computador e quanto tempo levou (em segundos) em 10 execuções diferentes; (iii) A Tabela 12 mostra os piores tempos encontrados para cada método usando diferentes máquinas *kernel* ($M = 70$, $LAG = 8$) para um processo inteiro (33.684 pontos de treinamento, 8.421 validação, e 168.413 teste).

O aumento do número de máquinas *kernel* afeta diretamente o tempo apresentado

Figura 24 – Comparação entre o RMSPE versus o tamanho do dicionário.



Fonte: Bezerra *et al.* (2021).

Tabela 12 – Tempo de execução (s)

	M2C KRLS	M2C KRLSW	M2C KRLSA	M2C KRLSAW	M3C KRLS	M3C KRLSW	M3C KRLSA	M3C KRLSAW
ALD-KRLS	205.1111	209.3029	206.0297	213.9421	301.7949	309.9622	307.4587	310.7236
SW-KRLS	139.3686	142.5556	139.9083	143.0825	245.3131	246.4561	242.0534	247.2417
KRLS-T	170.1855	172.3439	173.2392	177.2522	289.8388	294.6144	292.6144	297.4468
ALD-KRLS	100%	102.04%	100.45%	104.31%	147.14%	102.71%	101.88%	102.96%
SW-KRLS	100%	102.29%	100.39%	102.66%	176.02%	100.47%	98.67%	100.79%
KRLS-T	100%	101.79%	101.79%	104.15%	170.31%	101.65%	100.96%	102.62%

Fonte: Bezerra *et al.* (2021).

entre os grupos M2C e M3C. Independentemente do método usado, o tipo de máquina (ALD-KRLS, SW-KRLS, KRLS-T) tem um grande impacto no tempo de execução. Usando a média aritmética (MC-KRLS) como referência em todos os casos o uso da média ponderada (MC-KRLSW) trouxe um aumento de tempo, o uso do fator de envelhecimento (MC-KRLSA) aumentou o tempo de execução, mas não tanto quanto a versão ponderada, na verdade M3C-KRLSA (SW-KRLS) teve uma redução de 1,33% no tempo de execução. A combinação de ambos (MC-KRLSAW) apresenta o maior tempo de execução, resultado esperado pelo aumento das linhas de códigos e consequentemente de cálculos.

4.3 Conclusões

Neste capítulo os algoritmos propostos (*online* e *offline*) foram comparados com suas versões regulares, sendo apresentado o processo de maneira gradual, e didática, com o intuito de tornar o experimento reproduzível.

O algoritmo *offline* proposto, o PSO-AWF, aplicado ao treinamento de uma FTDNN, apresentou resultados melhores quando comparado à sua versão regular, mesmo com restrições como o tamanho do banco de dados utilizados no treinamento, e o aumento do intervalo de parâmetros utilizados pela ANN. O algoritmo *online* proposto, o MC-KRLS, é facilmente implementável e é capaz de atingir resultados melhores ou equiparáveis com suas versões clássicas.

5 CONSIDERAÇÕES FINAIS

Um método adaptativo, novo e competitivo, que pode ser aplicado em diversas ferramentas de inteligência computacional é apresentado neste trabalho. Este método, baseado no conceito de expectativa de vida, utiliza várias ferramentas que competem entre si, tendo seus resultados representados por dois fatores, de envelhecimento e enfraquecimento.

Neste estudo, da aplicação dos fatores foram proposto dois novos algoritmos: o PSO-AWF e o MC-KRLS. O PSO-AWF é caracterizado por atribuir a cada partícula do enxame características de envelhecimento e enfraquecimento. A expectativa de vida é ajustada de acordo com a aptidão de cada partícula.

Utilizando os conjuntos de dados BJ, MG, e LGD, a proposta do PSO-AWF foi mostrar que supera o BP, e o PSO, em termos de MSE e tempo de execução, ao ser utilizado para treinar uma FTDNN, com escolha dos parâmetros através da TE. O número de neurônios da camada escondida TE FTDNN varia para cada conjunto de dados. Utilizando o conjunto de dados WS, o PSO-AWF foi utilizado para treinar a FTDNN, e escolher dois parâmetros desta ANN.

Com os conjuntos de dados de referência, BJ, MG, e LGD, por se tratarem de conjuntos com poucos pontos, é normalmente fácil para as partículas melhorar a qualidade do enxame. Neste caso, as partículas têm uma expectativa de vida longa e o comportamento de procura do PSO-AWF é muito semelhante ao do PSO original. Por conseguinte, a característica de convergência rápida do PSO original pode ser preservada. Por outro lado, no conjunto de dados WS, que diferente dos outros conjuntos tem uma grande quantidade de pontos, uma vez que a partícula atinge um ótimo local, esta não consegue melhorar a qualidade do enxame e cessa rapidamente devido aos novos fatores introduzidos.

Para o PSO-AWF independente do tamanho do conjunto de dados novas partículas surgem para substituir as antigas e trazer diversidade. Portanto, o PSO-AWF tem a capacidade de escapar de ótimos locais e evitar a convergência prematura. A diferença entre o PSO e o PSO-AWF é o quão rápido isso acontece e está diretamente relacionada com as características do conjunto de dados. A eficácia do PSO-AWF FTDNN foi sempre superior, independente do uso de qualquer um dos métodos de treinamento (BP, DE e PSO) utilizado pelo TE FTDNN. O custo computacional do PSO-AWF FTDNN (TC3) é também melhor do que o do TE FTDNN (TC2), aproximadamente 59% em relação ao TE PSO FTDNN e 18% ao TE PSO-AWF FTDNN. O PSO-AWF obteve o melhor resultado e esteve presente em outras 6 das 8 melhores soluções.

Cinco destas soluções usaram o PSO-AWF como método de treinamento e escolha do valor do NNHL e LAG, outras duas o PSO-AWF foi usado como método de treinamento e a escolha do valor do NNHL e LAG feita através da TE.

Estudar os efeitos do envelhecimento sobre problemas de aprendizado de comitê de máquinas é importante. Como estes problemas requerem a manutenção de várias soluções ao mesmo tempo, incorporar o mecanismo de envelhecimento com técnicas de multipopulação, ou de conservação de espécies, seria útil.

O MC-KRLS pode ser usado em qualquer máquina *kernel*. O método permite usar várias máquinas relacionadas entre si pelo tamanho dos seus dicionários. Um fator de adaptação competitiva é adicionado quando todos eles atingem o seu tamanho máximo do dicionário, o pior *kernel* é então excluído. O algoritmo proposto cria novas matrizes *kernel* enquanto o processo estiver acontecendo, podendo assim, identificar mudanças na série temporal, evitando perda abrupta de informação, comum em métodos de rastreamento, e mantém o número de exemplos controlados por causa da seleção adaptativa de *kernels* ativos. Desta forma, o tamanho do *kernel* é fixo em vez de limitado. O método proposto foi aplicado a um conjunto de dois anos de dados de geração eólica de 23 parques eólicos na Austrália com capacidades entre 30 MW e 420 MW.

Dos 23 parques disponíveis 8 foram utilizadas no TC4 e TC5, sendo observado que o uso de máquinas com múltiplos *kernel* desafiantes tornou possível o uso de dicionários menores, obtendo-se melhores resultados, e reduzindo a probabilidade do *over-fitting*. A susceptibilidade dos modelos ao número de LAG também foi reduzida. Foram estudados os *trade-offs* entre o desempenho, RMSPE, e a complexidade de vários estado da arte de algoritmos *kernel*. Foram destacadas nesta pesquisa as vantagens e desvantagens de cada algoritmo em diferentes cenários. Para o horizonte de previsão de 5 minutos, em 75% dos modelos ALD MC-KRLS obteve melhores resultados do que os ALD-KRLS MIMO.

Para investigar ainda mais o desempenho dos modelos propostos, além do ALD-KRLS, duas máquinas *kernel*, SW-KRLS e KRLS-T, utilizando o mesmo conjunto de dados e metodologia foram avaliadas neste estudo. Mesmo com máquinas que têm ‘dispositivos de rastreamento’, um cenário semelhante foi encontrado. Com as SW MC-KRLS melhores resultados foram obtidos quando comparados às máquinas MIMO. O T MC-KRLS obteve resultados próximos aos modelos SISO e MIMO. De acordo com as Tabelas 11 e 12, o uso da expectativa de vida nas máquinas *kernel* geraram resultados melhores em 83,33% dos melhores resultados obtidos com acréscimos de tempo variando entre 2,66% e 4,31% do tempo da máquina

MC-KRLS, um caso inclusive apresenta uma redução no tempo de 1,33% (SW M3C-KRLSA) com resultado melhor.

A máquina do *kernel* escolhida impacta muito no tempo de computação. Observe que o tempo de cálculo pode ser uma questão durante o período de treinamento, validação e teste, mas durante a operação normal, o pior resultado para o preditor $t + 5$ minutos foi calculado em 1,47 milissegundos.

Observou-se que os algoritmos propostos trazem resultados positivos para a restrição dos dados armazenados pelos modelos, ao problema de *over-fitting*, e como esperado, não há um único algoritmo com melhor desempenho para todos os cenários. Em vez disso, a escolha ótima do algoritmo depende do intervalo alvo do RMSPE, dos recursos computacionais disponíveis, e do conjunto de dados.

5.1 Pesquisas futuras

Este trabalho foi motivado pela necessidade de produzir previsões precisas de curtíssimo prazo para um ou vários parques eólicos. Trabalhos futuros se concentrarão em:

- testar o mecanismo de envelhecimento em outras técnicas de computação evolutiva. Além disso, o uso de outros conjuntos de dados e funções iria fornecer mais informações sobre o comportamento dos fatores de envelhecimento e enfraquecimento; e
- estender esta abordagem a outras máquinas *kernel*; e desenvolvimento de outros modelos adaptativos, levando em consideração outros resultados de similaridade além do AE e MSE.
- definir um procedimento objetivo para garantir a qualidade dos dados históricos.

REFERÊNCIAS

- ACIKGOZ, H.; YILDIZ, C.; SEKKELI, M. An extreme learning machine based very short-term wind power forecasting method for complex terrain. **Energy Sources, Part A: Recovery, Utilization, and Environmental Effects**, London, United Kingdom, v. 42, n. 22, p. 2715–2730, 2020.
- ACKERMANN, T. *et al.* **Wind power in power systems**. Hoboken, NJ, USA: Wiley Online Library, 2005.
- AEMO. **Australian Energy Market Operator**. 2021. <http://www.aemo.com.au/>, Acesso em: 18 de maio de 2021.
- ALEXIADIS, M.; DOKOPOULOS, P.; SAHSAMANOGLU, H.; MANOUSARIDIS, I. Short-term forecasting of wind speed and related electrical power. **Solar Energy**, New York, NY, USA, v. 63, n. 1, p. 61–68, 1998.
- AMARI, S. A Theory of Adaptive Pattern Classifiers. **IEEE Transactions on Electronic Computers**, New York, NY, USA, EC-16, n. 3, p. 299–307, 1967.
- AMJADY, N.; KEYNIA, F.; ZAREIPOUR, H. Short-term wind power forecasting using ridgelet neural network. **Electric Power Systems Research**, New York, NY, USA, v. 81, n. 12, p. 2099–2107, 2011.
- ANDRADE, J. R.; BESSA, R. J. Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions. **IEEE Transactions on Sustainable Energy**, New York, NY, USA, v. 8, n. 4, p. 1571–1580, 2017.
- ARAGHI, A. R.; RIAHY, G.; CARLSON, O.; GROS, S. Enhancing the net energy of wind turbine using wind prediction and economic NMPC with high-accuracy nonlinear WT models. **Renewable Energy**, New York, NY, USA, v. 151, n. 0, p. 750–763, 2020.
- ARCE, F.; ZAMORA, E.; SOSSA, H.; BARRÓN, R. Dendrite morphological neural networks trained by differential evolution. In: **Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI)**. Athens, Greece: IEEE, 2016. p. 1–8.
- BEHESHTI, Z.; SHAMSUDDIN, S. M.; SULAIMAN, S. Fusion Global-Local-Topology Particle Swarm Optimization for Global Optimization Problems. **Mathematical Problems in Engineering**, London, United Kingdom, v. 2014, n. 0, p. 1–19, 2014.
- BERGH, F. Van den; ENGELBRECHT, A. P. Cooperative learning in neural networks using particle swarm optimizers. **South African Computer Journal**, Pretoria, Gauteng, South Africa, v. 26, n. 0, p. 84–90, 2000.
- BERGH, F. Van den; ENGELBRECHT, A. P. A cooperative approach to particle swarm optimization. **Evolutionary Computation, IEEE Transactions on**, New York, NY, USA, v. 8, n. 3, p. 225–239, 2004.
- BERGMEIR, C.; TRIGUERO, I.; MOLINA, D.; AZNARTE, J. L.; BENITEZ, J. M. Time Series Modeling and Forecasting Using Memetic Algorithms for Regime-Switching Models. **IEEE Transactions on Neural Networks and Learning Systems**, New York, NY, USA, v. 23, n. 11, p. 1841–1847, 2012.

- BESSA, R. J.; MIRANDA, V.; BOTTERUD, A.; WANG, J.; CONSTANTINESCU, E. M. Time Adaptive Conditional Kernel Density Estimation for Wind Power Forecasting. **IEEE Transactions on Sustainable Energy**, New York, NY, USA, v. 3, n. 4, p. 660–669, 2012.
- BETZ, A. Schraubenpropeller mit geringstem energieverlust. mit einem zusatz von l. prandtl. **Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse**, Göttingen, Germany, v. 1919, n. 0, p. 193–217, 1919.
- BEYER, H.; DEGNER, T.; HAUSMANN, J.; HOFFMANN, M.; RUJAN, P. Short term prediction of wind speed and power output of a wind turbine with neural networks. In: **Proceedings of the 1994 5th European Wind Energy Association Conference And Exhibition (EWEA)**. Thessaloniki, Greece: EWEA, 1994. p. 349–352.
- BEZERRA, E. C. **MC-KRLS collection of codes**. 2020. <https://github.com/erickbezerra/MC-KRLS>, Acesso em: 25 de agosto de 2020.
- BEZERRA, E. C.; LEÃO, R. P. S.; BRAGA, A. P. d. S. A Self-Adaptive Approach for Particle Swarm Optimization Applied to Wind Speed Forecasting. **Journal of Control, Automation and Electrical Systems**, Berlin/Heidelberg, Germany, v. 28, n. 6, p. 785–795, 2017.
- BEZERRA, E. C.; PINSON, P.; LEAO, R. P. S.; BRAGA, A. P. S. A Self-Adaptive Multikernel Machine Based on Recursive Least-Squares Applied to Very Short-Term Wind Power Forecasting. **IEEE Access**, New York, NY, USA, v. 9, n. 0, p. 104761–104772, 2021.
- BHASKAR, M.; JAIN, A.; SRINATH, N. V. Wind speed forecasting: Present status. In: **Proceedings of the 2010 International Conference on Power System Technology (POWERCON)**. Zhejiang, China: IEEE, 2010. p. 1–6.
- BLONBOU, R. Very short-term wind power forecasting with neural networks and adaptive Bayesian learning. **Renewable Energy**, New York, NY, USA, v. 36, n. 3, p. 1118–1124, 2011.
- BLUDSZUWEIT, H.; DOMINGUEZ-NAVARRO, J.; LLOMBART, A. Statistical Analysis of Wind Power Forecast Error. **IEEE Transactions on Power Systems**, New York, NY, USA, v. 23, n. 3, p. 983–991, 2008.
- BOLOIX-TORTOSA, R.; MURILLO-FUENTES, J. J.; SANTOS, I.; PEREZ-CRUZ, F. Widely Linear Complex-Valued Kernel Methods for Regression. **IEEE Transactions on Signal Processing**, New York, NY, USA, v. 65, n. 19, p. 5240–5248, 2017.
- BOUZGOU, H.; BENOUDJIT, N. Multiple architecture system for wind speed prediction. **Applied Energy**, New York, NY, USA, v. 88, n. 7, p. 2463–2471, 2011.
- BOX, G. E.; JENKINS, G. M. **Time series analysis forecasting and control**. San Francisco, CA, USA: [S. n.], 1970. <http://opac.inria.fr/record=b1108766>, Acesso em: 04 de fevereiro de 2022.
- BREMEN, L. von. Optimal linkage of nwp models with neural networks for offshore wind power predictions. In: **Proceedings of the 2006 Sixth International Workshop on Large-Scale Integration of Wind Power and Transmission Networks for Offshore Wind Farms**. Delft, Netherlands: Energynautics GmbH, 2006. p. 1–4.
- BREMNES, J. B. Probabilistic wind power forecasts using local quantile regression. **Wind Energy**, Hoboken, NJ, v. 7, n. 1, p. 47–54, 2004.

- BRYSON, A. E.; DENHAM, W. F. A steepest-ascent method for solving optimum programming problems. **Journal of Applied Mechanics**, New York, NY, USA, v. 29, n. 2, p. 247–257, 1962.
- CELIK, A. N.; KOLHE, M. Generalized feed-forward based method for wind energy prediction. **Applied Energy**, New York, NY, USA, v. 101, n. 0, p. 582–588, 2013.
- CHANG, W.-Y. A Literature Review of Wind Forecasting Methods. **Journal of Power and Energy Engineering**, Wuhan, China, v. 02, n. 04, p. 161–168, 2014.
- CHE, J.; WANG, J. Short-term load forecasting using a kernel-based support vector regression combination model. **Applied Energy**, New York, NY, USA, v. 132, n. 0, p. 602–609, 2014.
- CHEN, W.-N.; ZHANG, J.; LIN, Y.; CHEN, N.; ZHAN, Z.-H.; CHUNG, H. S.-H.; LI, Y.; SHI, Y.-H. Particle Swarm Optimization With an Aging Leader and Challengers. **IEEE Transactions on Evolutionary Computation**, New York, NY, USA, v. 17, n. 2, p. 241–258, 2013.
- CHONG, E. K.; ZAK, S. H. **An introduction to optimization**. Hoboken, NJ, USA: John Wiley & Sons, 2013.
- CLEMENTS, M. P. **Evaluating econometric forecasts of economic and financial variables**. Berlin/Heidelberg, Germany: Springer, 2005.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, Berlin/Heidelberg, Germany, v. 20, n. 3, p. 273–297, 1995.
- COSTA, A.; CRESPO, A.; NAVARRO, J.; LIZCANO, G.; MADSEN, H.; FEITOSA, E. A review on the young history of the wind power short-term prediction. **Renewable and Sustainable Energy Reviews**, New York, NY, USA, v. 12, n. 6, p. 1725–1744, 2008.
- CUTLER, N. **Wind energy forecasting issues paper - Version 12**. Sydney, Australia, 2006.
- DIEGO, I. M. de; MUÑOZ, A.; MOGUERZA, J. M. Methods for the combination of kernel matrices within a support vector framework. **Machine Learning**, Berlin/Heidelberg, Germany, v. 78, n. 1-2, p. 137–174, 2010.
- DONGMEI, Z.; YUCHEN, Z.; XU, Z. Research on wind power forecasting in wind farms. In: **Proceedings of the 2011 IEEE Power Engineering and Automation Conference (PEAM)**. Wuhan, China: IEEE, 2011. p. 175–178.
- DOUAK, F.; MELGANI, F.; BENOUDJIT, N. Kernel ridge regression with active learning for wind speed prediction. **Applied Energy**, New York, NY, USA, v. 103, n. 0, p. 328–340, 2013.
- DOWELL, J. **AEMO 5 minute wind power data**. Glasgow, United Kingdom: [S. n.], 2015. <https://pureportal.strath.ac.uk/en/datasets/australian-electricity-market-operator-aemo-5-minute-wind-power-d>, Acesso em: 21 de setembro de 2021.
- DOWELL, J.; PINSON, P. Very-Short-Term Probabilistic Wind Power Forecasts by Sparse Vector Autoregression. **IEEE Transactions on Smart Grid**, New York, NY, USA, v. 7, n. 2, p. 763–770, 2016.
- DOWELL, J.; WEISS, S.; INFELD, D. Kernel methods for short-term spatio-temporal wind prediction. In: **Proceedings of the 2015 IEEE Power & Energy Society General Meeting**. Denver, CO, USA: IEEE, 2015. p. 1–5.

DREYFUS, S. The computational solution of optimal control problems with time lag. **IEEE Transactions on Automatic Control**, New York, NY, USA, v. 18, n. 4, p. 383–385, 1973.

ELLOUMI, W.; BAKLOUTI, N.; ABRAHAM, A.; ALIMI, A. M. The multi-objective hybridization of particle swarm optimization and fuzzy ant colony optimization. **Journal of Intelligent & Fuzzy Systems**, Amsterdam, The Netherlands, v. 27, n. 1, p. 515–525, 2014.

ENGEL, Y.; MANNOR, S.; MEIR, R. The kernel recursive least-squares algorithm. **Signal Processing, IEEE Transactions on**, New York, NY, USA, v. 52, n. 8, p. 2275–2285, 2004.

ENGELBRECHT, A. P. **Computational intelligence: an introduction**. Hoboken, NJ, USA: John Wiley & Sons, 2007.

ESPINOZA, M.; SUYKENS, J. A.; BELMANS, R.; MOOR, B. D. Electric load forecasting: Using kernel-based modeling for nonlinear system identification. **IEEE Control Systems Magazine**, New York, NY, USA, v. 27, n. 5, p. 43–57, 2007.

FADARE, D. The application of artificial neural networks to mapping of wind speed profile for energy application in nigeria. **Applied Energy**, New York, NY, USA, v. 87, n. 3, p. 934–942, 2010.

FEI, S.-w.; WANG, M.-J.; MIAO, Y.-b.; TU, J.; LIU, C.-I. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. **Energy Conversion and Management**, New York, NY, USA, v. 50, n. 6, p. 1604–1609, 2009.

FOCKEN, U.; LANGE, M.; MÖNNICH, K.; WALDL, H.-P.; BEYER, H. G.; LUIG, A. Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects. **Journal of Wind Engineering and Industrial Aerodynamics**, New York, NY, USA, v. 90, n. 3, p. 231–246, 2002.

FOCKEN, U.; LANGE, M.; WALDL, H.-P. Reduction of wind power prediction error by spatial smoothing effects. In: **Proceedings of the 2001 European Wind Energy Conference and Exhibition (EWEC '01)**. Copenhagen, Denmark: EWEA, 2001. p. 822–825.

FOLEY, A.; LEAHY, P.; MCKEOGH, E. Wind power forecasting & prediction methods. In: **Proceedings of the 2010 9th International Conference on Environment and Electrical Engineering (EEEIC)**. Prague, Czech Republic: IEEE, 2010. p. 61–64.

FOLEY, A. M.; LEAHY, P. G.; MARVUGLIA, A.; MCKEOGH, E. J. Current methods and advances in forecasting of wind power generation. **Renewable Energy**, New York, NY, USA, v. 37, n. 1, p. 1–8, 2012.

FUKUDA, H.; TAMAKI, S.; NOKAMURA, M.; ONAGA, K.; NAGAI, H.; SHINJO, F.; ASATO, S. Pg1. 13 the development of a wind velocity prediction method based on a data-mining type auto-regressive model. In: **Proceedings of the 2001 European Wind Energy Conference and Exhibition (EWEC '01)**. Copenhagen, Denmark: EWEA, 2001. p. 741–744.

GALLEGO, C.; PINSON, P.; MADSEN, H.; COSTA, A.; CUERVA, A. Influence of local wind speed and direction on wind power dynamics—application to offshore very short-term forecasting. **Applied Energy**, New York, NY, USA, v. 88, n. 11, p. 4087–4096, 2011.

GARG, H. A hybrid pso-ga algorithm for constrained optimization problems. **Applied Mathematics and Computation**, New York, NY, USA, v. 274, n. 0, p. 292–305, 2016.

GAUTAM, C.; BALAJI, R.; SUDHARSAN, K.; TIWARI, A.; AHUJA, K. Localized multiple kernel learning for anomaly detection: one-class classification. **Knowledge-Based Systems**, New York, NY, USA, v. 165, n. 0, p. 241–252, 2019.

GAVRILOV, L. A.; GAVRILOVA, N. S. Evolutionary theories of aging and longevity. **The Scientific World Journal**, London, United Kingdom, v. 2, n. 0, p. 339–356, 2002.

GERSHENFELD, N. A.; WEIGEND, A. S. **The future of time series**. Palo Alto, CA, USA: Xerox Corporation, Palo Alto Research Center, 1993.

GIEBEL, G.; BROWNSWORD, R.; KARINIOTAKIS, G.; DENHARD, M.; DRAXL, C. **The state-of-the-art in short-term prediction of wind power: A literature overview**. Lyngby, Denmark, 2011.

GIEBEL, G.; KARINIOTAKIS, G. 3 - wind power forecasting—a review of the state of the art. In: **Renewable Energy Forecasting**. Sawston, Cambridge, England, United Kingdom: Woodhead Publishing, 2017. p. 59–109.

GOLDSMITH, T. C. Aging as an evolved characteristic—weismann’s theory reconsidered. **Medical hypotheses**, New York, NY, USA, v. 62, n. 2, p. 304–308, 2004.

GOLDSMITH, T. C. **The evolution of aging**. Annapolis, MD, USA: Azinet Press, 2006.

GÖNEN, M.; ALPAYDIN, E. Multiple kernel learning algorithms. **The Journal of Machine Learning Research**, Cambridge, MA, USA, v. 12, n. 64, p. 2211–2268, 2011.

GUDISE, V. G.; VENAYAGAMOORTHY, G. K. Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks. In: **Proceedings of the 2003 IEEE Swarm Intelligence Symposium (SIS’03)**. Indianapolis, IN, USA: IEEE, 2003. p. 110–117.

GüLCü Şaban; KODAZ, H. A novel parallel multi-swarm algorithm based on comprehensive learning particle swarm optimization. **Engineering Applications of Artificial Intelligence**, New York, NY, USA, v. 45, n. 0, p. 33–45, 2015.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation (3rd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2007.

HODGE, B.-M.; MILLIGAN, M. Wind power forecasting error distributions over multiple timescales. In: **Proceedings of the 2011 IEEE Power and Energy Society General Meeting**. Detroit, MI, USA: IEEE, 2011. p. 1–8.

HOLDEN, N.; FREITAS, A. A. A hybrid particle swarm/ant colony algorithm for the classification of hierarchical biological data. In: **Proceedings of the 2005 IEEE Swarm Intelligence Symposium (SIS 2005)**. Pasadena, CA, USA: IEEE, 2005. p. 100–107.

HOU, D.; KALNAY, E.; DROEGEMEIER, K. K. Objective verification of the samex ’98 ensemble forecasts. **Monthly Weather Review**, Boston, MA, USA, v. 129, n. 1, p. 73–91, 2001.

HU, X.; EBERHART, R. Multiobjective optimization using dynamic neighborhood particle swarm optimization. In: **Proceedings of the 2002 Congress on Evolutionary Computation (CEC’02) (Cat. No.02TH8600)**. Honolulu, HI, USA: IEEE, 2002. p. 1677–1681.

HUANG, G.-B.; WANG, D. H.; LAN, Y. Extreme learning machines: a survey. **International Journal of Machine Learning and Cybernetics**, Berlin/Heidelberg, Germany, v. 2, n. 2, p. 107–122, 2011.

HUEBNER, U.; ABRAHAM, N.; WEISS, C. Dimensions and entropies of chaotic intensity pulsations in a single-mode far-infrared NH_3 laser. **Phys Rev A Gen Phys**, Bethesda, MD, USA, v. 40, n. 11, p. 6354–6365, 1989.

HYNDMAN, R. J. **Time Series Data Library**. 2021. <https://robjhyndman.com/tsdl/>, Acesso em: 25 de agosto de 2021.

INC., T. M. **MATLAB version 8.4.0.150421 (R2014b)**. Natick, MA, USA: [S. n.], 2014.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An introduction to statistical learning**. Berlin/Heidelberg, Germany: Springer, 2013.

JIA, Y.; KWONG, S.; WU, W.; WANG, R.; GAO, W. Sparse Bayesian Learning-Based Kernel Poisson Regression. **IEEE Transactions on Cybernetics**, New York, NY, USA, v. 49, n. 1, p. 56–68, 2019.

JUANG, C.-F. A hybrid of genetic algorithm and particle swarm optimization for recurrent network design. **Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on**, New York, NY, USA, v. 34, n. 2, p. 997–1006, 2004.

JUBAN, J.; SIEBERT, N.; KARINIOTAKIS, G. N. Probabilistic short-term wind power forecasting for the optimal management of wind generation. In: **Proceedings of the 2007 IEEE Lausanne Power Tech**. Lausanne, Switzerland: IEEE, 2007. p. 683–688.

JUNG, J.; BROADWATER, R. P. Current status and future advances for wind speed and power forecasting. **Renewable and Sustainable Energy Reviews**, New York, NY, USA, v. 31, n. 0, p. 762–777, 2014.

JURSA, R. Variable selection for wind power prediction using particle swarm optimization. In: **Proceedings of the 9th annual conference on Genetic and evolutionary computation (GECCO'07)**. London, England: ACM, 2007. p. 2059–2065.

KABBARA, J.; PSAROMILIGKOS, I. N. Improving the tracking ability of kernels using kernel subspace pursuit. In: **Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. Florence, Italy: IEEE, 2014. p. 4543–4547.

KANNAO, R.; GUHA, P. Success based locally weighted multiple kernel combination. **Pattern Recognition**, New York, NY, USA, v. 68, n. 0, p. 38–51, 2017.

KARAKUŞ, O.; KURUOĞLU, E. E.; ALTINKAYA, M. A. One-day ahead wind speed/power prediction based on polynomial autoregressive model. **IET Renewable Power Generation**, London, United Kingdom, v. 11, n. 11, p. 1430–1439, 2017.

KARINIOTAKIS, G.; MARTI, I.; CASAS, D.; PINSON, P.; NIELSEN, T. S.; MADSEN, H.; GIEBEL, G.; USAOLA, J.; SANCHEZ, I.; PALOMARES, A. *et al.* What performance can be expected by short-term wind power prediction models depending on site characteristics. In: **Proceedings of the 2004 European Wind Energy Conference and Exhibition (EWEC '04)**. London, United Kingdom: EWEA, 2004. p. 2059–2065.

KARINIOTAKIS, G.; MAYER, D.; MOUSSAFIR, J.; CHEVALLAZ-PERRIER, R.; USAOLA, J.; SANCHEZ, I.; MARTI, I.; MADSEN, H.; NIELSEN, T. S.; LAC, C. *et al.* Anemos: development of a next generation wind power forecasting system for the large-scale integration of onshore & offshore wind farms. In: **Proceedings of the 2003 European Wind Energy Conference and Exhibition (EWECE '03)**. Madrid, Spain: EWEA, 2003. p. 2059–2065.

KARINIOTAKIS, G.; PINSON, P.; SIEBERT, N.; GIEBEL, G.; BARTHELMIE, R. *et al.* The state of the art in short-term prediction of wind power-from an offshore perspective. In: **Proceedings of the 2004 SeaTech Week - Ocean Energy Conference ADEME-IFREMER**. Brest, France: IEEE, 2004. p. 20–21.

KARINIOTAKIS, G.; STAVRAKAKIS, G.; NOGARET, E. Wind power forecasting using advanced neural networks models. **Energy conversion, iee transactions on**, New York, NY, USA, v. 11, n. 4, p. 762–767, 1996.

KELLEY, H. J. Gradient Theory of Optimal Flight Paths. **ARS Journal**, Reston, VA, USA, v. 30, n. 10, p. 947–954, 1960.

KENNEDY, J.; MENDES, R. Population structure and particle swarm performance. In: **Proceedings of the 2002 Congress on Evolutionary Computation (CEC'02) (Cat. No.02TH8600)**. Honolulu, HI, USA: IEEE, 2002. p. 1671–1676.

KHORRAMDEL, B.; CHUNG, C.; SAFARI, N.; PRICE, G. A fuzzy adaptive probabilistic wind power prediction framework using diffusion kernel density estimators. **IEEE Transactions on Power Systems**, New York, NY, USA, v. 33, n. 6, p. 7109–7121, 2018.

KOU, P.; GAO, F.; GUAN, X.; WU, J. Prediction intervals for wind power forecasting: Using sparse warped gaussian process. In: **Proceedings of the 2012 IEEE Power and Energy Society General Meeting**. San Diego, CA, USA: IEEE, 2012. p. 1–8.

KRETZSCHMAR, R.; ECKERT, P.; CATTANI, D.; EGGIMANN, F. Neural network classifiers for local wind prediction. **Journal of Applied Meteorology**, Boston, MA, USA, v. 43, n. 5, p. 727–738, 2004.

KUH, A.; MANDIC, D. Applications of complex augmented kernels to wind profile prediction. In: **Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)**. Taipei, Taiwan: IEEE, 2009. p. 3581–3584.

Kumari, P.; Wadhvani, R. Wind power prediction using klms algorithm. In: **Proceedings of the 2018 International Conference on Inventive Research in Computing Applications (ICIRCA)**. Coimbatore, India: IEEE, 2018. p. 154–161.

KWON, S.-D. Uncertainty analysis of wind energy potential assessment. **Applied Energy**, New York, NY, USA, v. 87, n. 3, p. 856–865, 2010.

LANDBERG, L. A mathematical look at a physical power prediction model. **Wind Energy**, Hoboken, NJ, USA, v. 1, n. 1, p. 23–28, 1999.

LANDBERG, L. Short-term prediction of the power production from wind farms. **Journal of Wind Engineering and Industrial Aerodynamics**, New York, NY, USA, v. 80, n. 1, p. 207–220, 1999.

LANDBERG, L. Short-term prediction of local wind conditions. **Journal of Wind Engineering and Industrial Aerodynamics**, New York, NY, USA, v. 89, n. 3, p. 235–245, 2001.

LANGE, M. **Analysis of the uncertainty of wind power predictions**. Tese (Doutorado) – Universität Oldenburg, Oldenburg, Germany, 2003. <https://d-nb.info/969985789/34>, Acesso em: 22 de novembro de 2022.

LANGE, M. On the uncertainty of wind power predictions-analysis of the forecast accuracy and statistical distribution of errors. **Journal of Solar Energy Engineering**, New York, NY, USA, v. 127, n. 2, p. 177–184, 2005.

LANGE, M.; FOCKEN, U. **Physical approach to short-term wind power prediction**. Berlin/Heidelberg, Germany: Springer, 2006.

LAPEDES, A.; FARBER, R. **LAUR872662 Nonlinear signal processing using neural networks: prediction and system modelling**. New Mexico, USA, 1987.

LEE, J.; ZHAO, F. **Global Wind Report 2019**. Brussels, Belgium, 2020.

LEI, M.; SHIYAN, L.; CHUANWEN, J.; HONGLING, L.; YAN, Z. A review on the forecasting of wind speed and generated power. **Renewable and Sustainable Energy Reviews**, New York, NY, USA, v. 13, n. 4, p. 915–920, 2009.

Li-Chiu Chang; Pin-An Chen; CHANG, F.-J. Reinforced Two-Step-Ahead Weight Adjustment Technique for Online Training of Recurrent Neural Networks. **IEEE Transactions on Neural Networks and Learning Systems**, New York, NY, USA, v. 23, n. 8, p. 1269–1278, 2012.

LI, G.; SHI, J.; ZHOU, J. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. **Renewable Energy**, New York, NY, USA, v. 36, n. 1, p. 352–359, 2011.

LI, N.; HE, F.; MA, W. Wind power prediction based on extreme learning machine with kernel mean p-power error loss. **Energies**, Basel, Switzerland, v. 12, n. 4, p. 673, 2019.

LIANG, J. J.; QIN, A. K.; SUGANTHAN, P. N.; BASKAR, S. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. **Evolutionary Computation, IEEE Transactions on**, New York, NY, USA, v. 10, n. 3, p. 281–295, 2006.

LIANG, J.-J.; SUGANTHAN, P. N. Dynamic multi-swarm particle swarm optimizer with local search. In: **Proceedings of the 2005 IEEE Congress on Evolutionary Computation**. Edinburgh, UK: IEEE, 2005. p. 522–528.

LIAO, Z.; GAO, L.; ZHOU, T.; FAN, X.; ZHANG, Y.; WU, J. An oil painters recognition method based on cluster multiple kernel learning algorithm. **IEEE Access**, New York, NY, USA, v. 7, n. 0, p. 26842–26854, 2019.

LIM, W. H.; ISA, N. A. M. Particle swarm optimization with increasing topology connectivity. **Engineering Applications of Artificial Intelligence**, New York, NY, USA, v. 27, n. 0, p. 80–102, 2014.

LIMA, D. K. S. d. **Modelagem de alta resolução para estimativa do potencial eólico offshore da costa do Ceará**. Tese (Doutorado) – Universidade Federal do Ceará, Fortaleza, Ceará, Brasil, 2015. <https://repositorio.ufc.br/handle/riufc/16338>, Acesso em: 22 de novembro de 2022.

- LIN, L.; ERIKSSON, J.; VIHRIALA, H.; SODERLUND, L. Predicting wind behavior with neural networks. In: **Proceedings of the 1996 European Wind Energy Conference and Exhibition (EWEC '96)**. Goeteborg, Sweden: EWEA, 1996. p. 655–658.
- LIU, F.; LI, R.; DREGLEA, A. Wind speed and power ultra short-term robust forecasting based on takagi–sugeno fuzzy model. **Energies**, Basel, Switzerland, v. 12, n. 18, p. 3551, 2019.
- LIU, H.; CHEN, C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. **Applied Energy**, New York, NY, USA, v. 249, n. 0, p. 392–408, 2019.
- LIU, H.; TIAN, H.-Q.; CHEN, C.; LI, Y.-f. A hybrid statistical method to predict wind speed and wind power. **Renewable Energy**, New York, NY, USA, v. 35, n. 8, p. 1857–1861, 2010.
- LIU, R.; LI, J.; FAN, J.; MU, C.; JIAO, L. A coevolutionary technique based on multi-swarm particle swarm optimization for dynamic multi-objective optimization. **European Journal of Operational Research**, New York, NY, USA, v. 261, n. 3, p. 1028 – 1051, 2017.
- LIU, T.; GU, Y. Multiple kernel learning for hyperspectral image classification. In: **Hyperspectral Image Analysis**. Berlin/Heidelberg, Germany: Springer, 2020. p. 259–293.
- LIU, Z.; JIANG, P.; ZHANG, L.; NIU, X. A combined forecasting model for time series: Application to short-term wind speed forecasting. **Applied Energy**, New York, NY, USA, v. 259, n. 0, p. 114137, 2020.
- LV, P.; YUE, L. Short-term wind speed forecasting based on non-stationary time series analysis and arch model. In: **Proceedings of the 2011 International Conference on Multimedia Technology (ICMT)**. Hangzhou, China: IEEE, 2011. p. 2549–2553.
- LYDIA, M.; KUMAR, S. S.; SELVAKUMAR, A. I.; KUMAR, G. E. P. Linear and non-linear autoregressive models for short-term wind speed forecasting. **Energy Conversion and Management**, New York, NY, USA, v. 112, n. 0, p. 115–124, 2016.
- MACKEY, M. C.; GLASS, L. Oscillation and chaos in physiological control systems. **Science**, Washington, DC, USA, v. 197, n. 4300, p. 287–289, 1977.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting methods and applications**. Hoboken, NJ, USA: John Wiley & Sons, 2008.
- MAO, B.; XIE, Z.; WANG, Y.; HANDROOS, H.; WU, H.; SHI, S. A hybrid differential evolution and particle swarm optimization algorithm for numerical kinematics solution of remote maintenance manipulators. **Fusion Engineering and Design**, New York, NY, USA, v. 124, n. 0, p. 587–590, 2017.
- MAO, M.; LING, J.; CHANG, L.; HATZIARGYRIOU, N. D.; ZHANG, J.; DING, Y. A Novel Short-Term Wind Speed Prediction Based on MFEC. **IEEE Journal of Emerging and Selected Topics in Power Electronics**, New York, NY, USA, v. 4, n. 4, p. 1206–1216, 2016.
- MARTÍ, I.; CABEZÓN, D.; VILLANUEVA, J.; SANISIDRO, M. J.; LOUREIRO, Y.; CANTERO, E.; SANZ, J. Localpred and regiopred. advanced tools for wind energy prediction in complex terrain. In: **Proceedings of the 2003 European Wind Energy Conference and Exhibition (EWEC '03)**. Madrid, Spain: EWEA, 2003. p. 16–19.

MARTI, I.; KARINIOTAKIS, G.; PINSON, P.; SANCHEZ, I.; NIELSEN, T.; MADSEN, H.; GIEBEL, G.; USAOLA, J.; PALOMARES, A. M.; BROWNSWORD, R. *et al.* Evaluation of advanced wind power forecasting models—results of the anemos project. In: **Proceedings of the 2006 European Wind Energy Conference and Exhibition (EWEC '06)**. Athènes, Greece: EWEA, 2006. p. 1–9.

MATHEWS, V. J.; SICURANZA, G. L. **Polynomial signal processing**. Hoboken, NJ, USA: Wiley-Interscience, 2000.

MCCARTHY, E.; LANDBERG, L.; ZACK, J.; YEN, D.; MCGOWIN, C. **California wind energy forecasting system development and testing. Phase 1: Initial testing**. Palo Alto, CA, USA, 2003.

MESSNER, J. W.; PINSON, P.; BROWELL, J.; BJERREGÅRD, M. B.; SCHICKER, I. Evaluation of wind power forecasts—an up-to-date view. **Wind Energy**, Hoboken, NJ, USA, v. 23, n. 6, p. 1461–1481, 2020.

MILLIGAN, M.; SCHWARTZ, M.; WAN, Y.-h. **Statistical wind power forecasting for US wind farms**. Golden, CO, USA, 2004.

MILLIGAN, M. R.; MILLER, A. H.; CHAPMAN, F. **Estimating the economic value of wind forecasting to utilities**. Golden, CO, USA, 1995.

MIRANDA, M. S.; DUNN, R. W.; LI, F.; SHADDICK, G.; BELL, K. R. Bayesian inferencing for wind resource characterisation. In: **Proceedings of the 2006 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)**. Stockholm, Sweden: IEEE, 2006. p. 1–6.

MISHRA, S.; DASH, P. Short-term prediction of wind power using a hybrid pseudo-inverse legendre neural network and adaptive firefly algorithm. **Neural Computing and Applications**, Berlin/Heidelberg, Germany, v. 31, n. 7, p. 2243–2268, 2019.

MOGUERZA, J. M.; MUÑOZ, A.; DIEGO, I. M. Improving support vector classification via the combination of multiple sources of information. In: **Structural, Syntactic, and Statistical Pattern Recognition**. Berlin/Heidelberg, Germany: Springer, 2004. p. 592–600.

MONTEIRO, C.; BESSA, R.; MIRANDA, V.; BOTTERUD, A.; WANG, J.; CONZELMANN, G. **Wind power forecasting: state-of-the-art 2009**. [*S. l.*], 2009.

MORALES, J. M.; MINGUEZ, R.; CONEJO, A. J. A methodology to generate statistically dependent wind speed scenarios. **Applied Energy**, New York, NY, USA, v. 87, n. 3, p. 843–855, 2010.

MORE, A.; DEO, M. Forecasting wind with neural networks. **Marine structures**, New York, NY, USA, v. 16, n. 1, p. 35–49, 2003.

NIELSEN, H. A.; MADSEN, H.; NIELSEN, T. S. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. **Wind Energy**, Hoboken, NJ, USA, v. 9, n. 1-2, p. 95–108, 2006.

NIELSEN, H. A.; NIELSEN, T. S.; MADSEN, H.; GIEBEL, G.; BADGER, J.; LANDBERG, L.; SATTLER, K.; VOULUND, L.; TOFTING, J. From wind ensembles to probabilistic information about future wind power production—results from an actual application. In: **Proceedings of**

the **2006 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)**. Stockholm, Sweden: IEEE, 2006. p. 1–8.

OKUMUS, I.; DINLER, A. Current status of wind energy forecasting and a hybrid method for hourly predictions. **Energy Conversion and Management**, New York, NY, USA, v. 123, n. 0, p. 362–371, 2016.

PAVLIDIS, P.; WESTON, J.; CAI, J.; GRUNDY, W. N. Gene functional classification from heterogeneous data. In: **Proceedings of 2001 RECOMB01: The Fifth Annual International Conference on Computational Molecular Biology**. Montreal, Quebec, Canada: ACM, 2001. p. 249–255.

PAWAR, A.; JAPE, V.; MATHEW, S. Wind power forecasting using support vector machine model in rstudio. In: **Cognitive Informatics and Soft Computing**. Berlin/Heidelberg, Germany: Springer, 2019. p. 289–298.

PENG, J.; ZHU, X.; WANG, Y.; AN, L.; SHEN, D. Structured sparsity regularized multiple kernel learning for alzheimer’s disease diagnosis. **Pattern recognition**, New York, NY, USA, v. 88, n. 0, p. 370–382, 2019.

PINSON, P. **Estimation of the uncertainty in wind power forecasting**. Tese (Doutorado) – École Nationale Supérieure des Mines de Paris, Paris, France, 2006. <https://pastel.archives-ouvertes.fr/pastel-00002187>, Acesso em: 22 de novembro de 2022.

PINSON, P.; CHEVALLIER, C.; KARINIOTAKIS, G. *et al.* Optimizing benefits from wind power participation in electricity markets using advanced tools for wind power forecasting and uncertainty assessment. In: **Proceedings of the 2004 European Wind Energy Conference and Exhibition (EWEC ’04)**. London, United Kingdom: EWEA, 2004. p. 1–8.

PINSON, P.; CHEVALLIER, C.; KARINIOTAKIS, G. N. Trading wind generation from short-term probabilistic forecasts of wind power. **Power Systems, IEEE Transactions on**, New York, NY, USA, v. 22, n. 3, p. 1148–1156, 2007.

PINSON, P.; KARINIOTAKIS, G. On-line assessment of prediction risk for wind power production forecasts. **Wind Energy**, Hoboken, NJ, USA, v. 7, n. 2, p. 119–132, 2004.

PINSON, P.; KARINIOTAKIS, G.; NIELSEN, H. A.; NIELSEN, T. S.; MADSEN, H. Properties of quantile and interval forecasts of wind generation and their evaluation. In: **Proceedings of the 2006 European Wind Energy Conference and Exhibition (EWEC ’06)**. Athènes, Greece: EWEA, 2006. p. 128–133.

PINSON, P.; MADSEN, H.; NIELSEN, H. A.; PAPAETHYMIU, G.; KLÖCKL, B. From probabilistic forecasts to statistical scenarios of short-term wind power production. **Wind energy**, Hoboken, NJ, USA, v. 12, n. 1, p. 51–62, 2009.

PINSON, P.; NIELSEN, H. A.; MADSEN, H.; KARINIOTAKIS, G. Skill forecasting from ensemble predictions of wind power. **Applied Energy**, New York, NY, USA, v. 86, n. 7, p. 1326–1334, 2009.

PINSON, P.; NIELSEN, H. A.; MÖLLER, J. K.; MADSEN, H.; KARINIOTAKIS, G. N. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. **Wind Energy**, Hoboken, NJ, USA, v. 10, n. 6, p. 497–516, 2007.

PINSON, P.; SIEBERT, N.; KARINIOTAKIS, G. Forecasting of regional wind generation by a dynamic fuzzy-neural networks based upscaling approach. In: **Proceedings of the 2003 European Wind Energy Conference and Exhibition (EWEC '03)**. Athènes, Greece: EWEA, 2003. p. 1–5.

PRINCIPE, J. C.; EULIANO, N. R.; LEFEBVRE, W. C. **Neural and adaptive systems: fundamentals through simulations with CD-ROM**. Hoboken, NJ, USA: John Wiley & Sons, 1999.

QIAN, Z.; PEI, Y.; ZAREIPOUR, H.; CHEN, N. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. **Applied Energy**, New York, NY, USA, v. 235, n. 0, p. 939–953, 2019.

QIU, S.; LANE, T. A framework for multiple kernel support vector regression and its applications to sirna efficacy prediction. **Computational Biology and Bioinformatics, IEEE/ACM Transactions on**, New York, NY, USA, v. 6, n. 2, p. 190–199, 2009.

QU, Z.; MAO, W.; ZHANG, K.; ZHANG, W.; LI, Z. Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network. **Renewable energy**, New York, NY, USA, v. 133, n. 0, p. 919–929, 2019.

RAHIMI, A.; GÖNEN, M. A multitask multiple kernel learning formulation for discriminating early-and late-stage cancers. **Bioinformatics**, Oxford, United Kingdom, v. 36, n. 12, p. 3766–3772, 2020.

RANKINEN, R. **FORECASTING WIND ENERGY FOR A DATA CENTER**. Dissertação (Mestrado) – University of Oulu, Oulu, Finland, 2021. <http://jultika.oulu.fi/Record/nbnfioulu-202106188578>, Acesso em: 22 de novembro de 2022.

RATHMANN, O. S.; HANSEN, B. O.; HANSEN, K. S.; MORTENSEN, N. G.; LEON, J. P. M. **E-0160 The Park2 Wake Model-Documentation and Validation**. Roskilde, Denmark, 2018.

RAZA, S.; MOKHLIS, H.; AROF, H.; NAIDU, K.; LAGHARI, J. A.; KHAIRUDDIN, A. S. M. Minimum-features-based ANN-PSO approach for islanding detection in distribution system. **IET Renewable Power Generation**, London, United Kingdom, v. 10, n. 9, p. 1255–1263, 2016.

REN, Z.; LI, H.; YANG, C.; SUN, Q. Multiple kernel subspace clustering with local structural graph and low-rank consensus kernel learning. **Knowledge-Based Systems**, New York, NY, USA, v. 188, n. 0, p. 105040, 2020.

RICHARD, C.; BERMUDEZ, J. C. M.; HONEINE, P. Online prediction of time series data with kernels. **Signal Processing, IEEE Transactions on**, New York, NY, USA, v. 57, n. 3, p. 1058–1067, 2009.

RICHARD, P. **The effects of integrating wind power on transmission system planning, reliability, and operations**. Schenectady, NY, USA, 2005.

ROUNGKVIST, J. S.; ENEVOLDSEN, P. Timescale classification in wind forecasting: A review of the state-of-the-art. **Journal of Forecasting**, Hoboken, NJ, USA, v. 39, n. 5, p. 757–768, 2020.

SAINI, N.; SAHA, S. Multi-objective optimization techniques: a survey of the state-of-the-art and applications. **The European Physical Journal Special Topics**, Berlin/Heidelberg, Germany, v. 230, n. 10, p. 2319–2335, 2021.

SANTHOSH, M.; VENKAIAH, C.; KUMAR, D. M. V. Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: A review. **Engineering Reports**, Hoboken, NJ, USA, v. 2, n. 6, p. 104761–104772, 2020.

SCHÖLKOPF, B.; SMOLA, A. J. **Learning with kernels: Support vector machines, regularization, optimization, and beyond**. Cambridge, MA, USA: MIT press, 2002.

SFETSOS, A. A comparison of various forecasting techniques applied to mean hourly wind speed time series. **Renewable energy**, New York, NY, USA, v. 21, n. 1, p. 23–35, 2000.

SFETSOS, A. A comprehensive overview of short term wind forecasting models based on time series analysis. In: **Soft Computing in Green and Renewable Energy Systems**. Berlin/Heidelberg, Germany: Springer, 2011. p. 97–116.

SHAWE-TAYLOR, J.; CRISTIANINI, N. **Kernel Methods for Pattern Analysis**. Cambridge, England: Cambridge University Press, 2004.

SHEKIN, D. J. **Handbook of parametric and nonparametric statistical procedures**. Boca Raton, FL, USA: Chapman & Hall/CRC, 2000.

SHI, Y.; EBERHART, R. A modified particle swarm optimizer. In: **Proceedings of the 1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)**. Anchorage, AK, USA: IEEE, 1998. p. 69–73.

SHI, Y.; EBERHART, R. C. Empirical study of particle swarm optimization. In: **Proceedings of the 1999 Congress on Evolutionary Computation (CEC99) (Cat. No. 99TH8406)**. Washington, DC, USA: IEEE, 1999. p. 1945–1950.

SHI, Y.; EBERHART, R. C. Fuzzy adaptive particle swarm optimization. In: **Proceedings of the 2001 Congress on Evolutionary Computation (Cat. No.01TH8546)**. Seoul, South Korea: IEEE, 2001. p. 101–106.

SIDERATOS, G.; HATZIARGYRIOU, N. D. An advanced statistical method for wind power forecasting. **Power Systems, IEEE Transactions on**, New York, NY, USA, v. 22, n. 1, p. 258–265, 2007.

SIEBERT, N. **Development of methods for regional wind power forecasting**. Tese (Doutorado) – École Nationale Supérieure des Mines de Paris, Paris, France, 2008. <https://tel.archives-ouvertes.fr/tel-00287551>, Acesso em: 22 de novembro de 2022.

SIM, S.-K.; MAASS, P.; LIND, P. G. Wind speed modeling by nested arima processes. **Energies**, Basel, Switzerland, v. 12, n. 1, p. 69, 2019.

SINGH, S.; MOHAPATRA, A. *et al.* Repeated wavelet transform based arima model for very short-term wind speed forecasting. **Renewable energy**, New York, NY, USA, v. 136, n. 0, p. 758–768, 2019.

SMOLA, A.; VAPNIK, V. Support vector regression machines. In: **Proceedings of the 1997 Conference on Neural Information Processing Systems (NeurIPS)**. Denver, CO, USA: NIPS, 1997. p. 155–161.

- SOMAN, S. S.; ZAREIPOUR, H.; MALIK, O.; MANDAL, P. A review of wind power and wind speed forecasting methods with different time horizons. In: **Proceedings of the 2010 North American Power Symposium (NAPS)**. Arlington, TX, USA: IEEE, 2010. p. 1–8.
- SONG. A new approach for wind speed prediction. **Wind Engineering**, Newbury Park, CA, USA, v. 24, n. 1, p. 35–47, 2000.
- SORENSEN, B.; MEIBOM, P. Can wind power be sold in a deregulated electricity market? In: **Proceedings of the 1999 European Wind Energy Conference and Exhibition (EWEC '99)**. Nice, France: EWEA, 1999. p. 375–378.
- SSEKULIMA, E. B.; ANWAR, M. B.; HINAI, A. A.; MOURSI, M. S. E. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review. **IET Renewable Power Generation**, London, United Kingdom, v. 10, n. 7, p. 885–989, 2016.
- STORN, R.; PRICE, K. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. **Journal of Global Optimization**, Berlin/Heidelberg, Germany, v. 11, n. 4, p. 341–359, 1997.
- SUGANTHAN, P. N. Particle swarm optimiser with neighbourhood operator. In: **Proceedings of the 1999 Congress on Evolutionary Computation (CEC99) (Cat. No. 99TH8406)**. Washington, DC, USA: IEEE, 1999. p. 1958–1962.
- TAMBKE, J.; BREMEN, L. V.; BARTHELMIE, R.; PALOMARES, A. M.; RANCHIN, T.; JUBAN, J.; KARINIOTAKIS, G.; BROWNSWORD, R.; WALDL, I. Short-term forecasting of offshore wind farm production—developments of the anemos project. In: **Proceedings of the 2006 European Wind Energy Conference and Exhibition (EWEC '06)**. Athènes, Greece: EWEA, 2006. p. 1–13.
- TAMBKE, J.; LANGE, M.; FOCKEN, U.; WOLFF, J.-O.; BYE, J. A. Forecasting offshore wind speeds above the north sea. **Wind Energy**, Hoboken, NJ, USA, v. 8, n. 1, p. 3–16, 2005.
- TASTU, J.; PINSON, P.; KOTWA, E.; MADSEN, H.; NIELSEN, H. A. Spatio-temporal analysis and modeling of short-term wind power forecast errors. **Wind Energy**, Hoboken, NJ, USA, v. 14, n. 1, p. 43–60, 2011.
- TASTU, J.; PINSON, P.; MADSEN, H. Multivariate conditional parametric models for a spatio-temporal analysis of short-term wind power forecast errors. In: **Proceedings of the 2010 European Wind Energy Conference and Exhibition (EWEC '10)**. Warsaw, Poland: EWEA, 2010. p. 77–81.
- TAYLOR, K. E. Summarizing multiple aspects of model performance in a single diagram. **Journal of Geophysical Research: Atmospheres**, Hoboken, NJ, USA, v. 106, n. D7, p. 7183–7192, 2001.
- THØGERSEN, M. L.; MOTTA, M.; SØRENSEN, T.; NIELSEN, P. Measure-correlate-predict methods: case studies and software implementation. In: **Proceedings of the 2007 European Wind Energy Conference and Exhibition (EWEC '07)**. Milan, Italy: EWEA, 2007. p. 1–10.
- TIAN, Z.; LI, S.; WANG, Y. A prediction approach using ensemble empirical mode decomposition-permutation entropy and regularized extreme learning machine for short-term wind speed. **Wind Energy**, Hoboken, NJ, USA, v. 23, n. 2, p. 177–206, 2020.

- TIAN, Z.; LI, S.; WANG, Y.; WANG, X. Wind power prediction method based on hybrid kernel function support vector machine. **Wind Engineering**, London, United Kingdom, v. 42, n. 3, p. 252–264, 2018.
- TILLET, J.; RAO, T.; SAHIN, F.; RAO, R. Darwinian particle swarm optimization. In: **Proceedings of the 2005 2nd Indian International Conference on Artificial Intelligence (IICAI-05)**. Pune, India: IICAI, 2005. p. 1474–1487.
- TOBAR, F. A.; Sun-Yuan Kung; MANDIC, D. P. Multikernel Least Mean Square Algorithm. **IEEE Transactions on Neural Networks and Learning Systems**, New York, NY, USA, v. 25, n. 2, p. 265–277, 2014.
- TUKEY, J. W. **Exploratory data analysis**. Boston, MA, USA: Reading Mass: Addison-Wesley, 1977.
- VAERENBERGH, S. V.; LÁZARO-GREDILLA, M.; SANTAMARÍA, I. Kernel recursive least-squares tracker for time-varying regression. **Neural Networks and Learning Systems, IEEE Transactions on**, New York, NY, USA, v. 23, n. 8, p. 1313–1326, 2012.
- VAERENBERGH, S. V.; SANTAMARÍA, I. A comparative study of kernel adaptive filtering algorithms. In: **Proceedings of the 2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)**. Napa, CA, USA: IEEE, 2013. p. 1–9.
- VAERENBERGH, S. V.; VIA, J.; SANTAMARÍA, I. A sliding-window kernel rls algorithm and its application to nonlinear channel identification. In: **Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (ICASSP 2006)**. Toulouse, France: IEEE, 2006. p. 1–4.
- VAPNIK, V. **Estimation of dependences based on empirical data**. Berlin/Heidelberg, Germany: Springer Science & Business Media, 2006.
- VAPNIK, V. **The nature of statistical learning theory**. Berlin/Heidelberg, Germany: Springer Science & Business Media, 2013.
- VAPNIK, V. N. An overview of statistical learning theory. **Neural Networks, IEEE Transactions on**, New York, NY, USA, v. 10, n. 5, p. 988–999, 1999.
- VARGAS, S. A.; ESTEVES, G. R. T.; MAÇAIRA, P. M.; BASTOS, B. Q.; OLIVEIRA, F. L. C.; SOUZA, R. C. Wind power generation: A review and a research agenda. **Journal of Cleaner Production**, New York, NY, USA, v. 218, n. 0, p. 850–870, 2019.
- WALLIS, K. F. Chi-squared tests of interval and density forecasts, and the bank of england's fan charts. **International Journal of Forecasting**, New York, NY, USA, v. 19, n. 2, p. 165–175, 2003.
- WANG, J.; ZHU, S.; ZHANG, W.; LU, H. Combined modeling for electric load forecasting with adaptive particle swarm optimization. **Energy**, New York, NY, USA, v. 35, n. 4, p. 1671–1678, 2010.
- WANG, M.-D.; QIU, Q.-R.; CUI, B.-W. Short-term wind speed forecasting combined time series method and arch model. In: **Proceedings of the 2012 International Conference on Machine Learning and Cybernetics (ICMLC)**. Xi'an, China: IEEE, 2012. p. 924–927.

- WANG, Q.; FU, G.; LI, L.; WANG, H.; LI, Y. Data-dependent multiple kernel learning algorithm based on soft-grouping. **Pattern Recognition Letters**, New York, NY, USA, v. 112, n. 0, p. 111–117, 2018.
- WANG, S.-X.; LI, M.; ZHAO, L.; JIN, C. Short-term wind power prediction based on improved small-world neural network. **Neural Computing and Applications**, Berlin/Heidelberg, Germany, v. 31, n. 7, p. 3173–3185, 2019.
- WANG, Z.; ZHU, Z.; LI, D. Collaborative and geometric multi-kernel learning for multi-class classification. **Pattern Recognition**, New York, NY, USA, v. 99, n. 0, p. 107050, 2020.
- WEI, C.; CHEN, J.; SONG, Z.; CHEN, C.-I. Development of Self-Learning Kernel Regression Models for Virtual Sensors on Nonlinear Processes. **IEEE Transactions on Automation Science and Engineering**, New York, NY, USA, v. 16, n. 1, p. 286–297, 2019.
- WEIGEND, A. S.; GERSHENFELD, N. A. Time series prediction: Forecasting the future and understanding the past. In: **Proceedings of the 1992 NATO Advanced Research Workshop on Comparative Time Series Analysis**. Santa Fe, NM, USA: American Statistical Association, 1994. p. 1149.
- WEISMANN, A. **Ueber die Dauer des Lebens**. Berlin/Heidelberg, Germany: G. Fischer, 1882.
- WILKINSON, J. **The algebraic eigenvalue problem**. New York, NY, USA: Springer-Verlag New York, 1971.
- WILLIAMS, D. R. G. H. R.; HINTON, G. Learning representations by back-propagating errors. **Nature**, London, United Kingdom, v. 323, n. 0, p. 533–536, 1986.
- WU, S.-H.; DU, K.-J.; ZHAN, Z.-H.; WANG, H.; ZHANG, J. Historical information-based differential evolution for dynamic optimization problem. In: **Proceedings of the 2021 IEEE Congress on Evolutionary Computation (CEC '21)**. Kraków, Poland: IEEE, 2021. p. 119–126.
- YAN, W. Toward automatic time-series forecasting using neural networks. **Neural Networks and Learning Systems, IEEE Transactions on**, New York, NY, USA, v. 23, n. 7, p. 1028–1039, 2012.
- YEH, W.-C. New parameter-free simplified swarm optimization for artificial neural network training and its application in the prediction of time series. **Neural Networks and Learning Systems, IEEE Transactions on**, New York, NY, USA, v. 24, n. 4, p. 661–665, 2013.
- YOUSUF, M. U.; AL-BAHADLY, I.; AVCI, E. Current Perspective on the Accuracy of Deterministic Wind Speed and Power Forecasting. **IEEE Access**, New York, NY, USA, v. 7, n. 0, p. 159547–159564, 2019.
- ZHANG, J.; CUI, M.; HODGE, B.-M.; FLORITA, A.; FREEDMAN, J. Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales. **Energy**, New York, NY, USA, v. 122, n. 0, p. 528–541, 2017.
- ZHANG, Q.; LAI, K. K.; NIU, D.; WANG, Q.; ZHANG, X. A fuzzy group forecasting model based on least squares support vector machine (ls-svm) for short-term wind power. **Energies**, Basel, Switzerland, v. 5, n. 9, p. 3329–3346, 2012.

ZHANG, W.-J.; XIE, X.-F. *et al.* Depso: hybrid particle swarm with differential evolution operator. In: **Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics (SMC'03) (Cat. No.03CH37483)**. Washington, DC, USA: IEEE, 2003. p. 3816–3821.

ZHANG, X.; ZHANG, J.; LI, Y.; ZHANG, R. Short-term forecasting of wind speed based on recursive least squares. In: **Proceedings of the 2011 International Conference on Electrical and Control Engineering (ICECE)**. Yichang, China: IEEE, 2011. p. 367–370.

ZHANG, Y.; WANG, S.; JI, G. A comprehensive survey on particle swarm optimization algorithm and its applications. **Mathematical Problems in Engineering**, Cairo, Egypt, v. 2015, n. 0, p. 1–38, 2015.

ZHAO, L.; YANG, Y. Pso-based single multiplicative neuron model for time series prediction. **Expert Systems with Applications**, New York, NY, USA, v. 36, n. 2, p. 2805–2812, 2009.

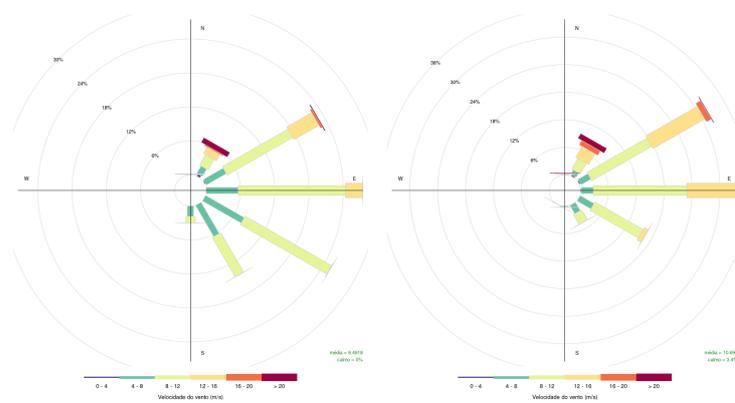
ZHAO, X.; WANG, S.; LI, T. Review of evaluation criteria and main methods of wind power forecasting. **Energy Procedia**, New York, NY, USA, v. 12, n. 0, p. 761–769, 2011.

ZHENG, H.; LI, H.; LU, X.; RUAN, T. A multiple kernel learning approach for air quality prediction. **Advances in Meteorology**, Cairo, Egypt, v. 2018, n. 0, p. 1–15, 2018.

ZHOU, P.; YE, F.; DU, L. Unsupervised robust multiple kernel learning via extracting local and global noises. **IEEE Access**, New York, NY, USA, v. 7, n. 0, p. 34451–34461, 2019.

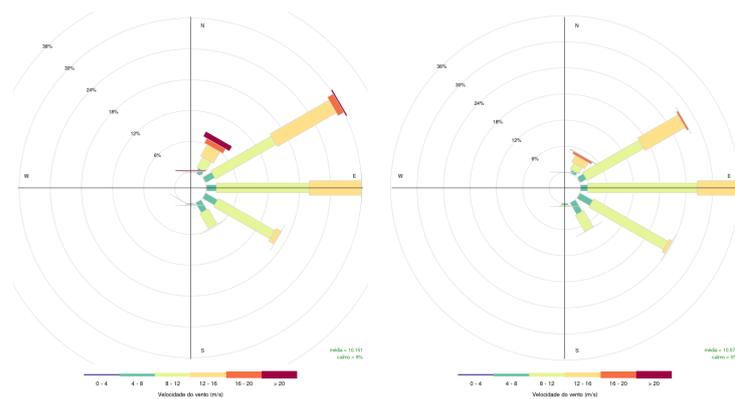
APÊNDICE A – ROSA DOS VENTOS DO CONJUNTO DE DADOS WS

Figura 25 – Rosa dos ventos dos anos.



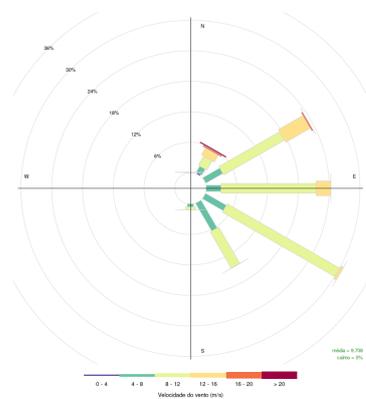
(a) 2009

(b) 2010



(c) 2011

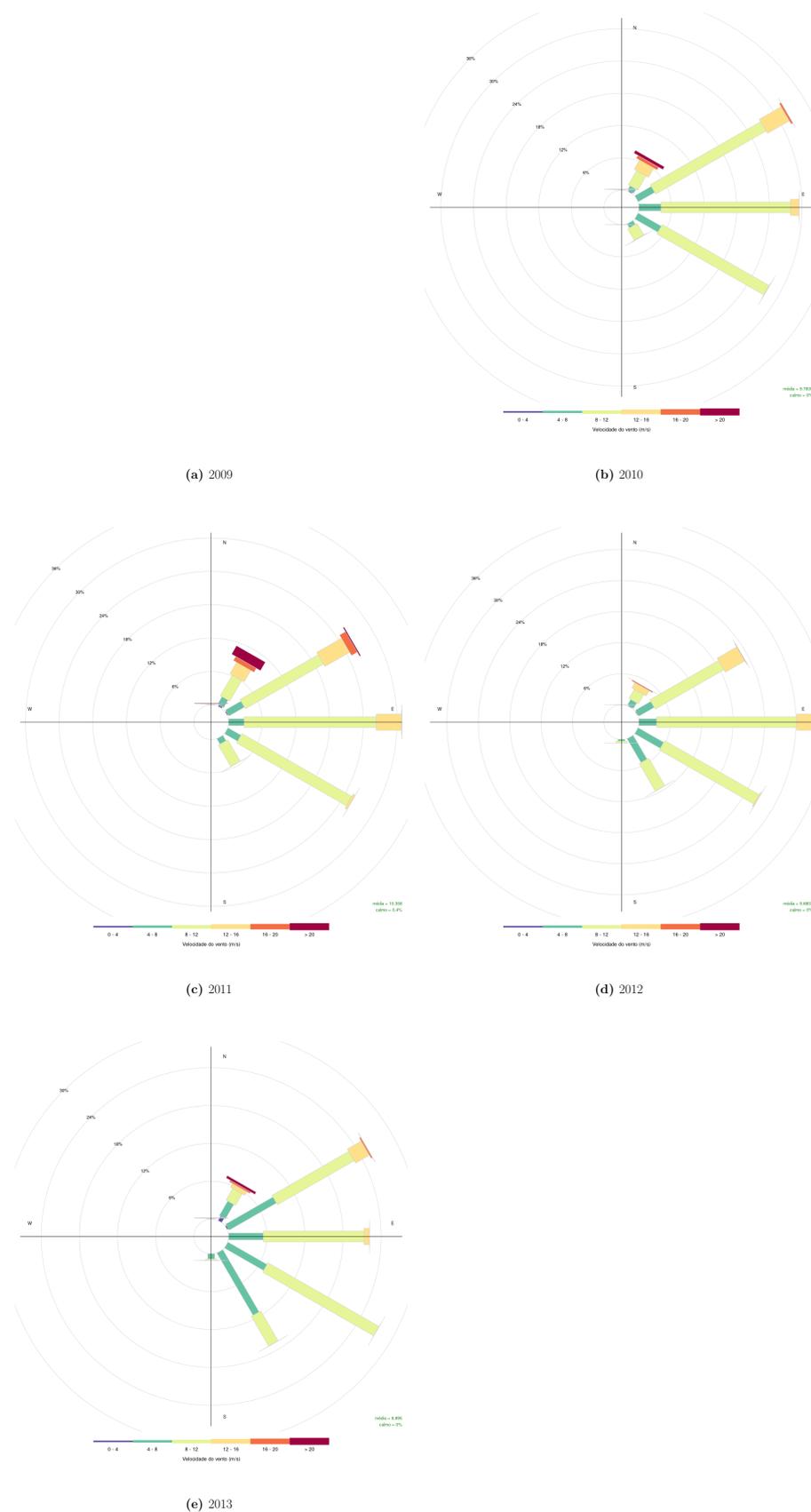
(d) 2012



(e) 2013

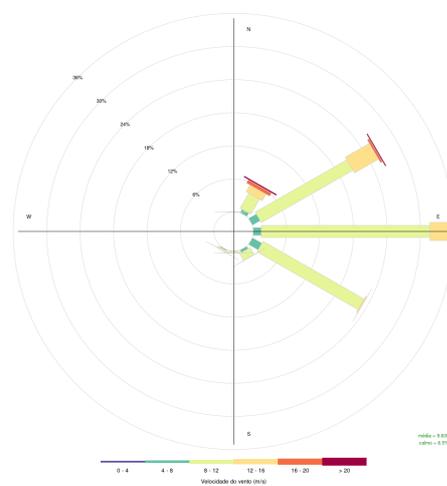
Fonte: Elaborado pelo autor (2021).

Figura 26 – Rosa dos ventos de janeiro.



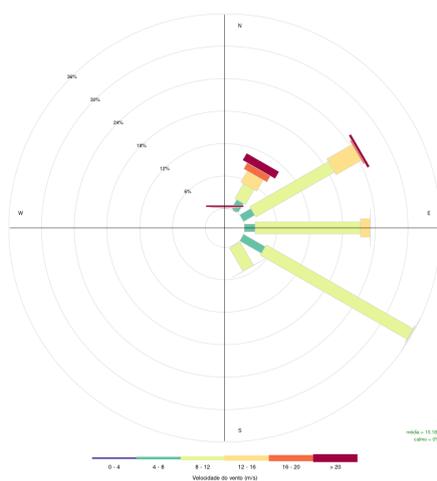
Fonte: Elaborado pelo autor (2021).

Figura 27 – Rosa dos ventos de fevereiro.

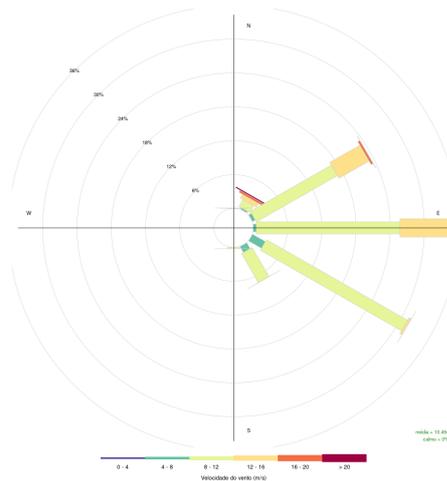


(a) 2009

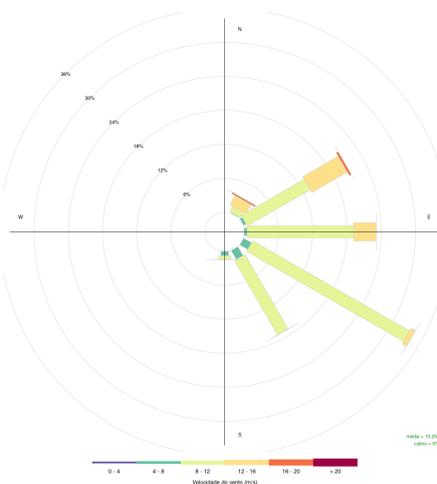
(b) 2010



(c) 2011



(d) 2012



(e) 2013

Fonte: Elaborado pelo autor (2021).

Figura 28 – Rosa dos ventos de março.

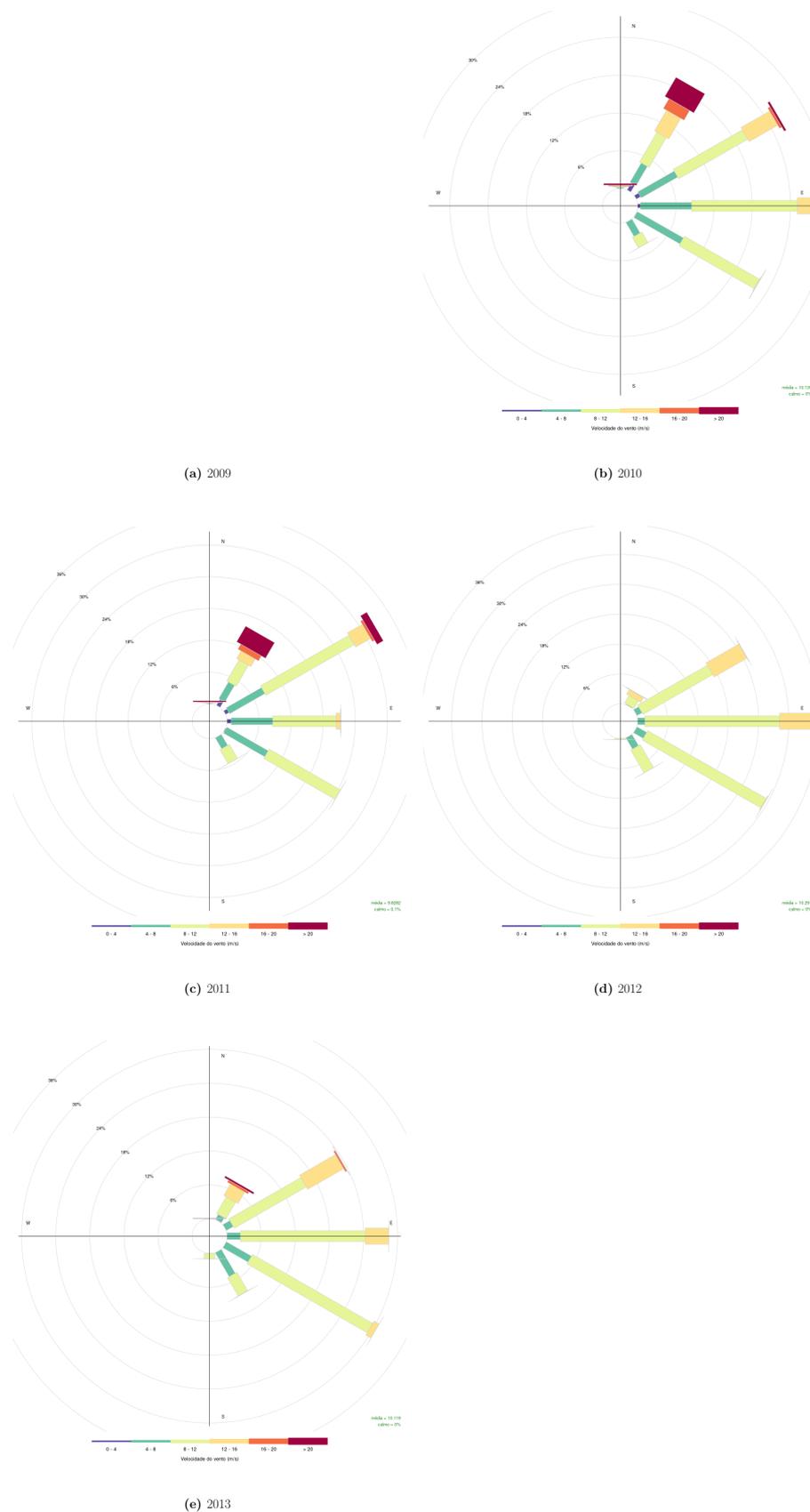
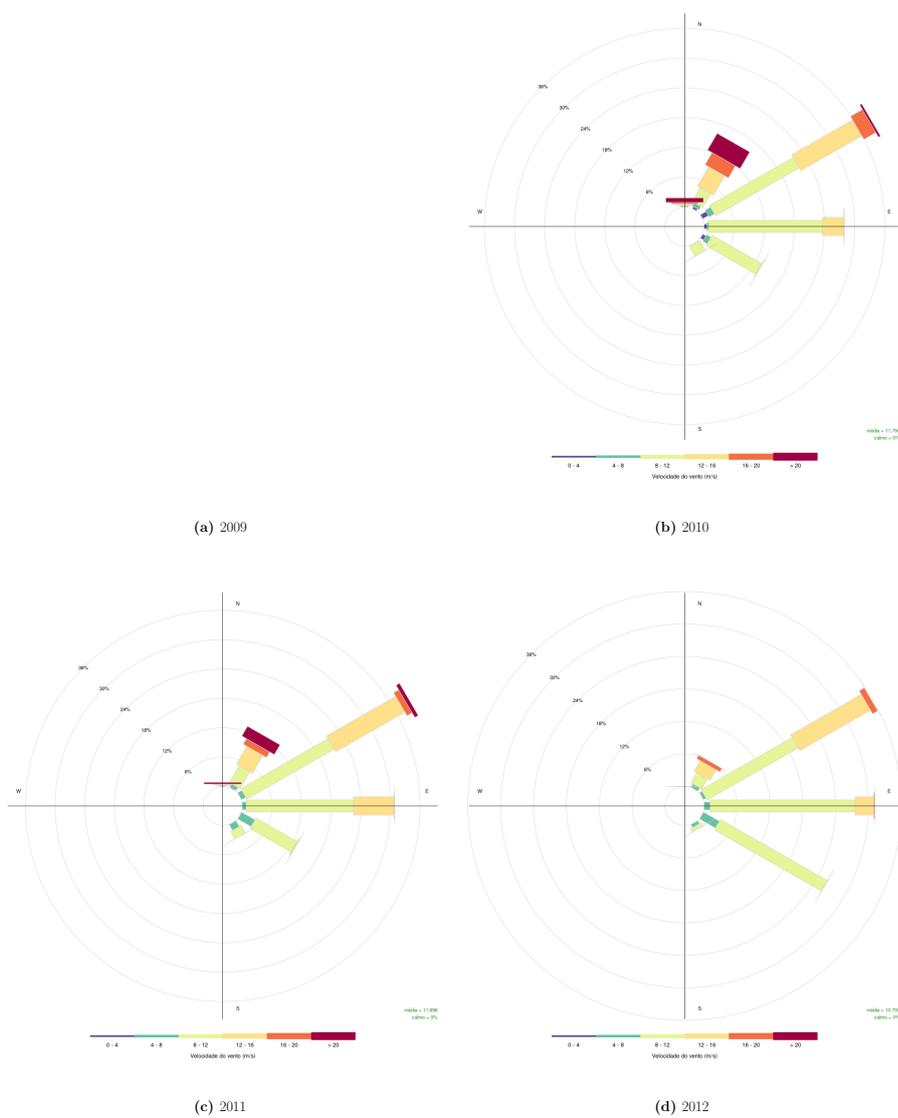


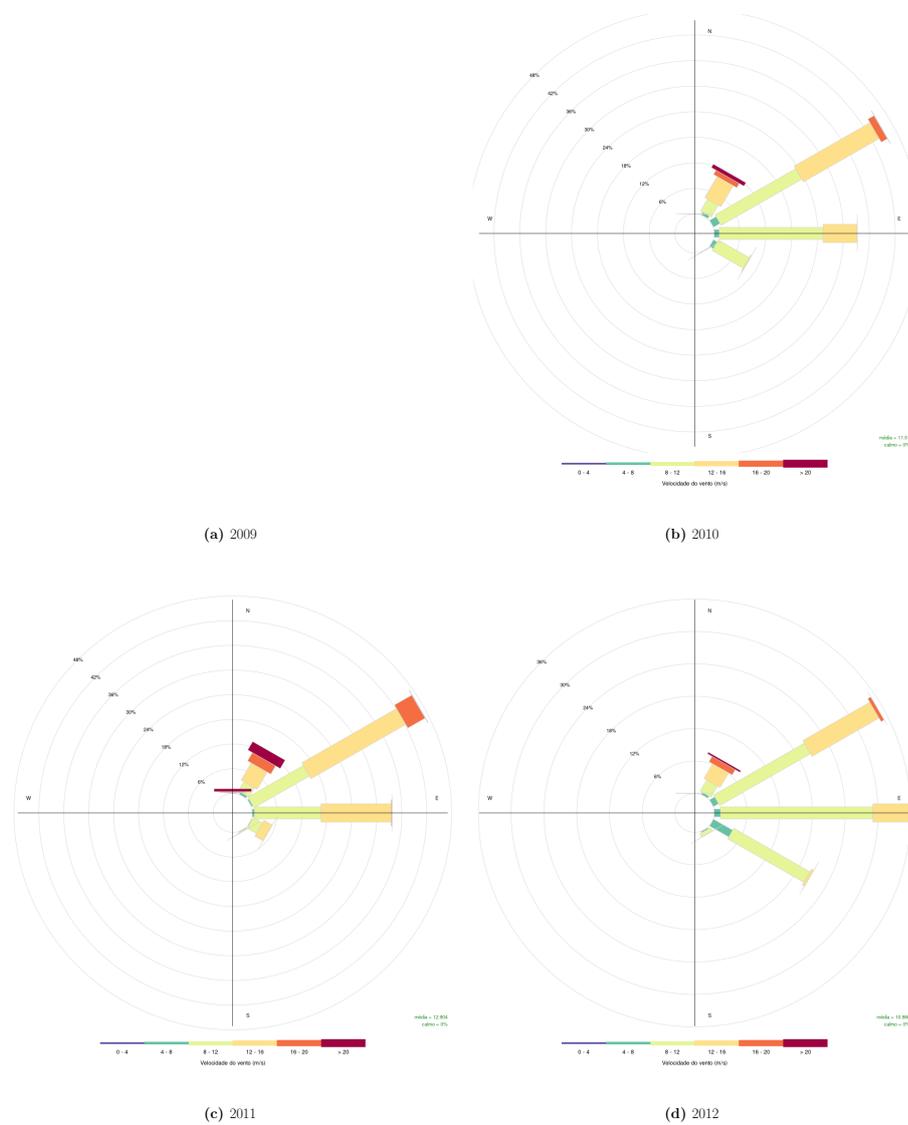
Figura 29 – Rosa dos ventos de abril.



(e) 2013

Fonte: Elaborado pelo autor (2021).

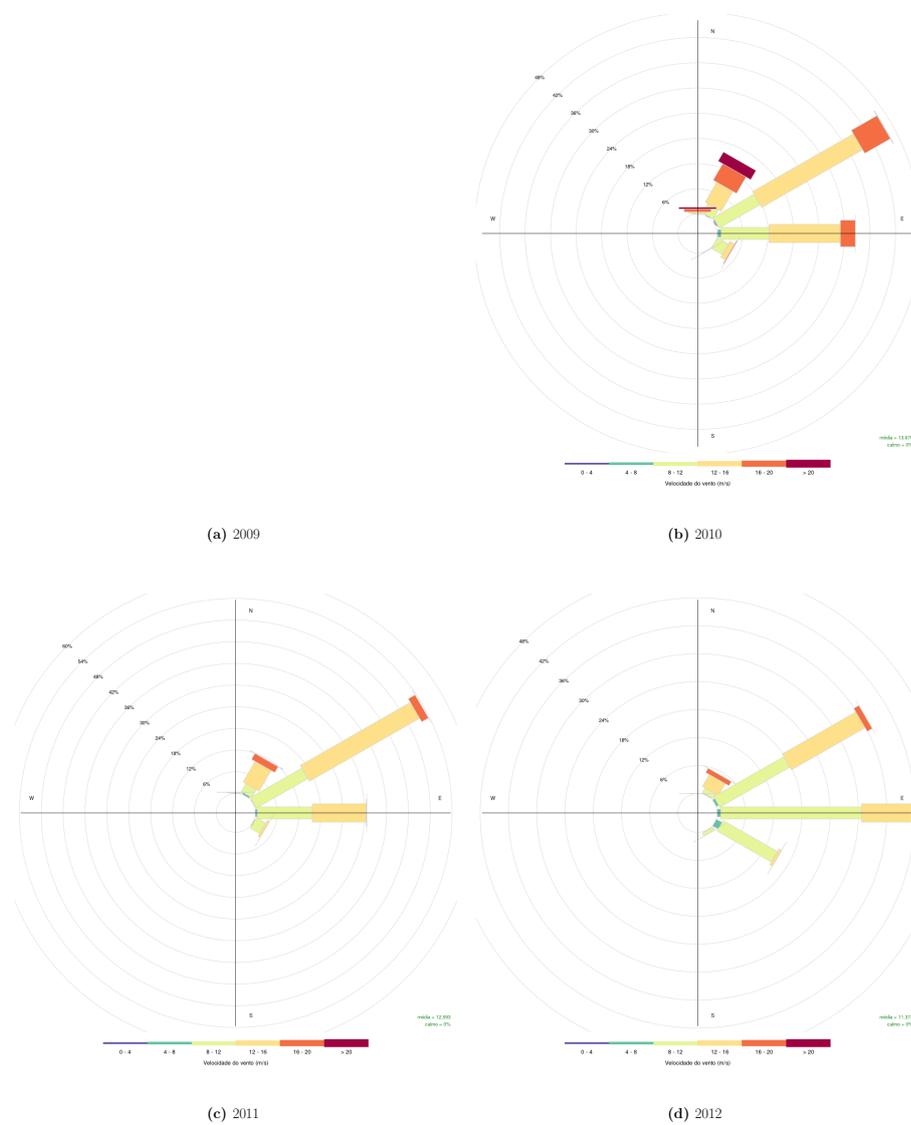
Figura 30 – Rosa dos ventos de maio.



(e) 2013

Fonte: Elaborado pelo autor (2021).

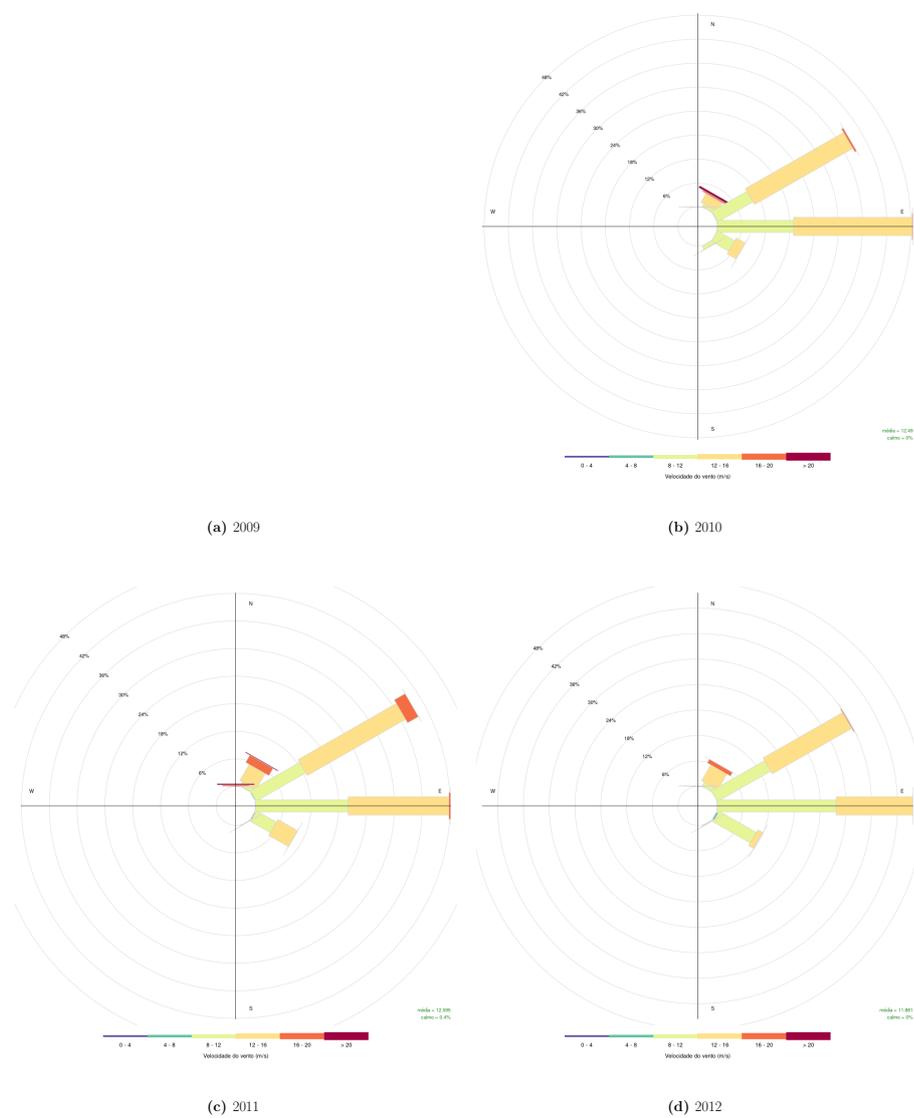
Figura 31 – Rosa dos ventos de junho.



(e) 2013

Fonte: Elaborado pelo autor (2021).

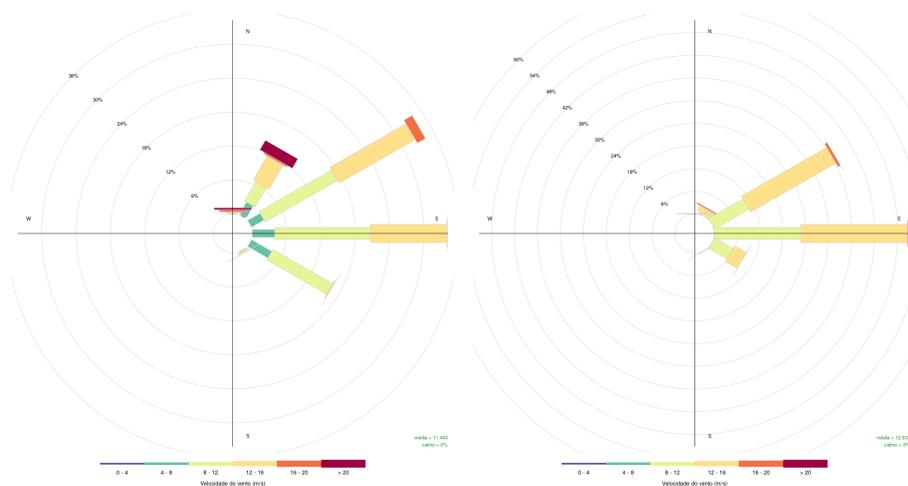
Figura 32 – Rosa dos ventos de julho.



(e) 2013

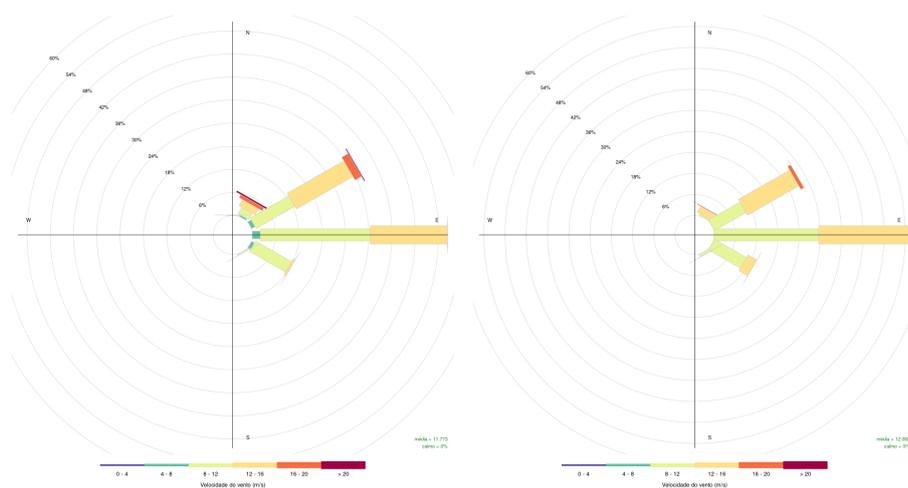
Fonte: Elaborado pelo autor (2021).

Figura 33 – Rosa dos ventos de agosto.



(a) 2009

(b) 2010



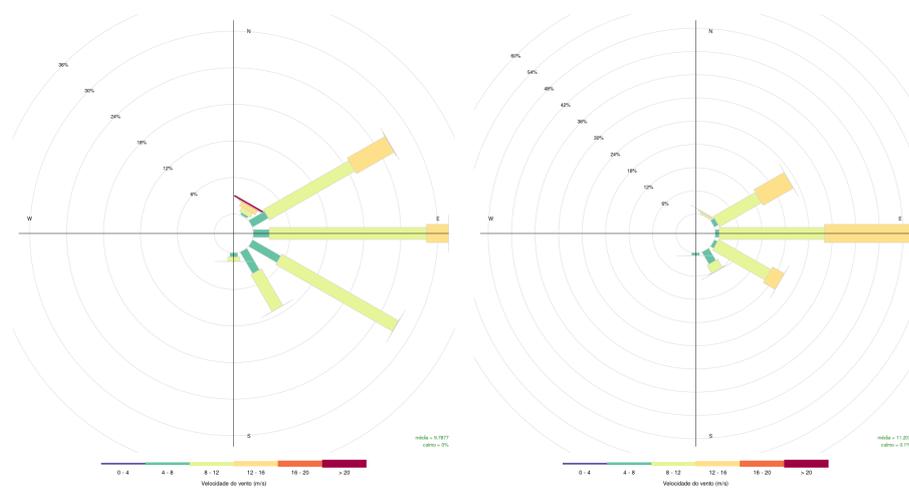
(c) 2011

(d) 2012

(e) 2013

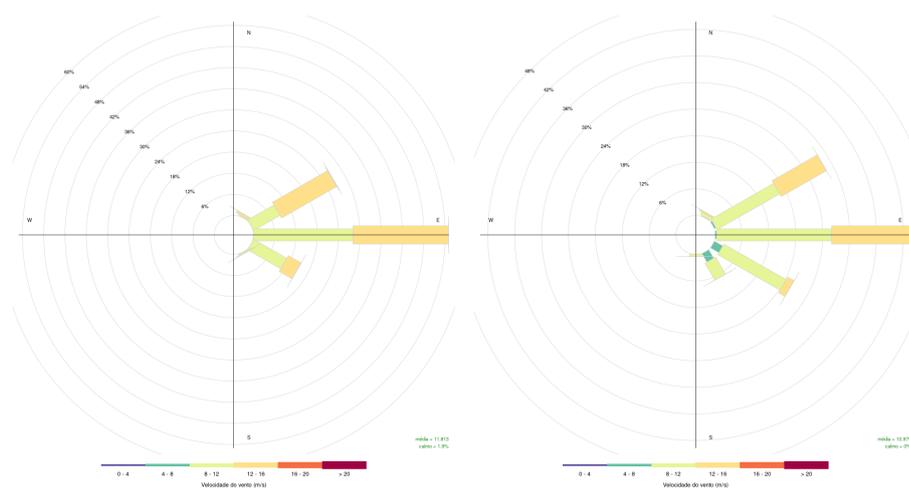
Fonte: Elaborado pelo autor (2021).

Figura 34 – Rosa dos ventos de setembro.



(a) 2009

(b) 2010



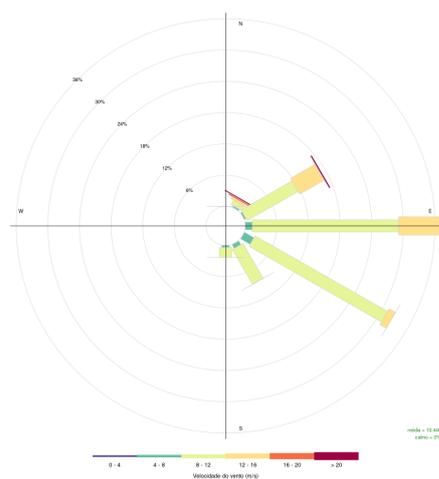
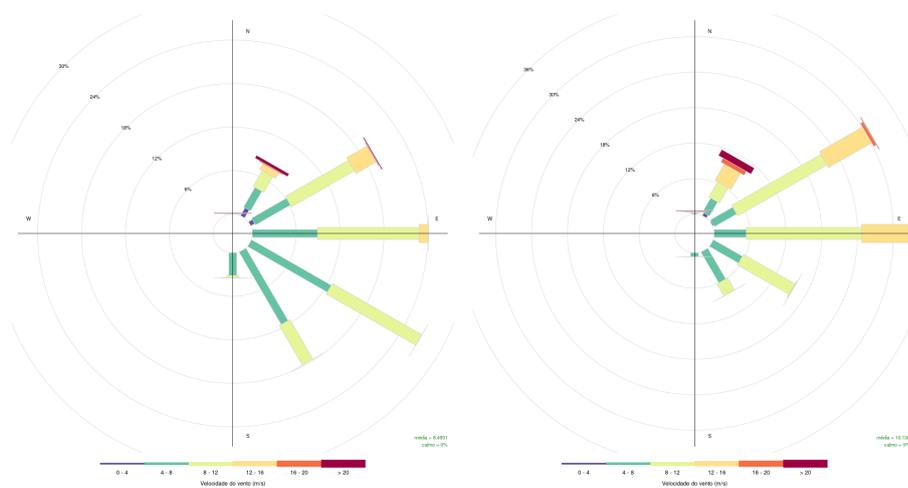
(c) 2011

(d) 2012

(e) 2013

Fonte: Elaborado pelo autor (2021).

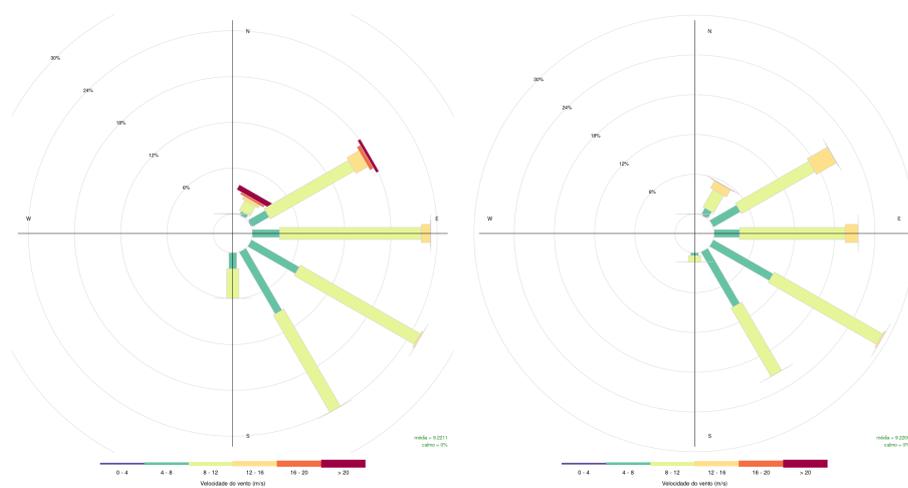
Figura 35 – Rosa dos ventos de outubro.



(e) 2013

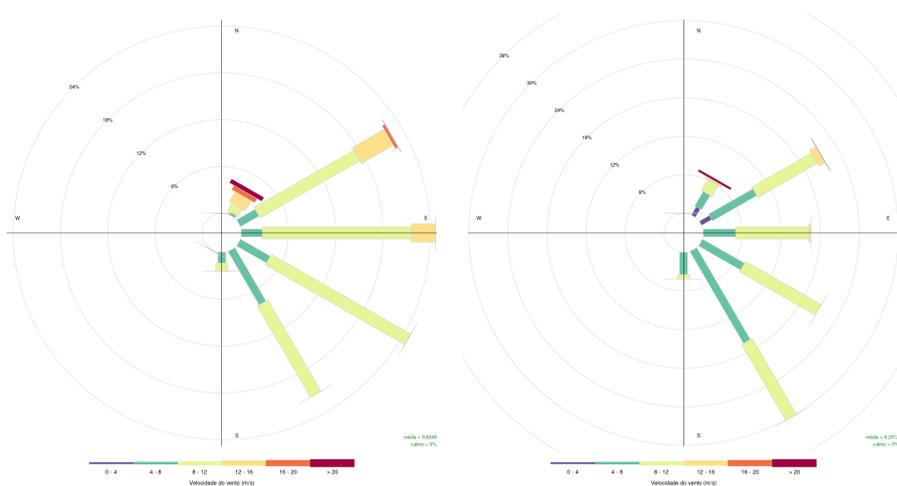
Fonte: Elaborado pelo autor (2021).

Figura 36 – Rosa dos ventos de novembro.



(a) 2009

(b) 2010



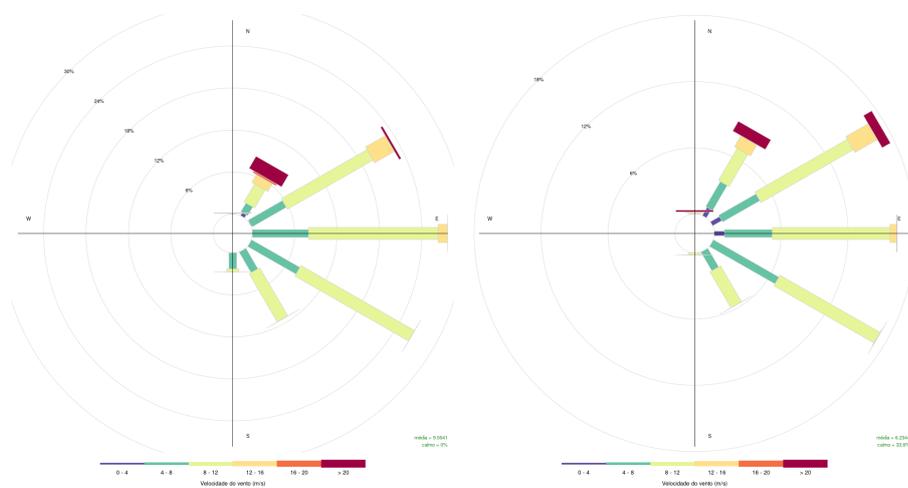
(c) 2011

(d) 2012

(e) 2013

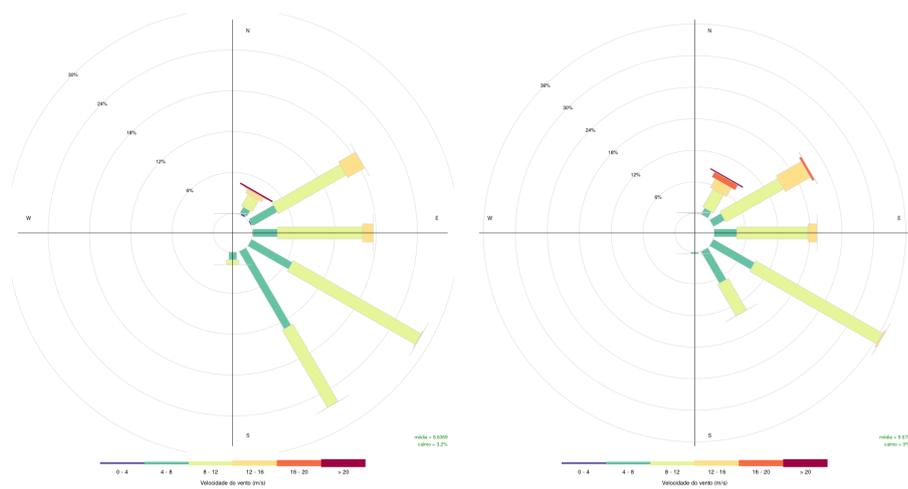
Fonte: Elaborado pelo autor (2021).

Figura 37 – Rosa dos ventos de dezembro.



(a) 2009

(b) 2010



(c) 2011

(d) 2012

(e) 2013

Fonte: Elaborado pelo autor (2021).