

**Taciana Araújo da Silva**

**Estudo do desempenho da combinação de preditores  
baseados em cópulas e máquinas de vetor de suporte para  
séries temporais úteis ao desenvolvimento sustentável**

**Recife**

**12 de maio de 2020**



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

**Estudo do desempenho da combinação de preditores baseados em cópulas e máquinas de vetor de suporte para séries temporais úteis ao desenvolvimento sustentável**

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 28/02/2020 pela comissão examinadora

**Área de concentração: Biometria e Estatística Aplicada**

**Orientador: Dr. Frank Sinatra Gomes da Silva**

**Coorientador: Dr. Paulo Renato Alves Firmino**

**Recife**

**12 de maio de 2020**

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema Integrado de Bibliotecas da UFRPE  
Biblioteca Central, Recife-PE, Brasil

S586e Silva, Taciana Araújo da  
Estudo do desempenho da combinação de preditores baseados em cópulas e máquinas de vetor de suporte para séries temporais úteis ao desenvolvimento sustentável / Taciana Araújo da Silva. – 2020.  
98 f. : il.

Orientador: Frank Sinatra Gomes da Silva.  
Coorientador: Paulo Renato Alves Firmino.  
Dissertação (Mestrado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, BR-PE, 2020.  
Inclui referências.

1. Análise séries temporais 2. Desenvolvimento sustentável  
3. Cópulas (Estatística matemática) 4. Campos vetoriais  
5. Modelagem I. Silva, Frank Sinatra Gomes da, orient. II. Firmino, Paulo Renato Alves, coorient. III. Título

CDD 519.5

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

**Estudo do desempenho da combinação de preditores baseados em cópulas e máquinas de vetor de suporte para séries temporais úteis ao desenvolvimento sustentável**

Taciana Araújo da Silva

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 28/02/2020 pela comissão examinadora

Orientador:

---

**Dr. Frank Sinatra Gomes da Silva**  
Orientador

Banca examinadora:

---

**Dr. Josimar Mendes de Vasconcelos**  
Universidade Federal Rural de  
Pernambuco

---

**Dr. Paulo Renato Alves Firmino**  
Universidade Federal do Cariri

---

**Dr. Paulo Salgado Gomes de Mattos**  
Neto  
Universidade Federal de Pernambuco

---

*Este trabalho é dedicado à minha mãe,  
ao meu pai (in memoriam), ao meu noivo,  
família, afilhados e amigos.*

# Agradecimentos

Agradeço primeiramente a Deus por me permitir trilhar os caminhos que me trouxeram até aqui, pela tranquilidade e sabedoria para concluir esta etapa.

Agradeço a minha mãe, Fatima, mulher forte, guerreira, que tudo fez por mim e que jamais me deixou desistir. A imensidão de seu coração e amor nunca serão mensuráveis.

Agradeço ao meu pai, Antônio (*in memoriam*), homem de uma sabedoria admirável e caráter inquestionável, por ser um exemplo de vida.

Agradeço ao meu noivo, LUSDÊNIO, a quem dedico muita admiração pelo caráter e empenho, por todo incentivo e apoio.

Agradeço a toda minha família, em especial minha avó Nelza e tia Luzanira, por serem fonte de conforto nas horas mais difíceis. Aos amigos, em especial Kerolly e Fábio, por todo apoio e ajuda, sem os quais impossibilitariam minha permanência no curso. Aos meus afilhados Iavlla e Luís Pedro, pelas alegrias e a imensidão de amor. Aos colegas de curso, pela troca de experiências.

Agradeço aos professores do ensino fundamental e médio, em especial aos que mostraram encanto pela profissão, desempenhando tão bem o seu papel, se fazendo fonte de inspiração, por todo conhecimento recebido. Aos professores da Universidade Regional do Cariri, em especial aos professores Kátia Pires e Pedro Lima, que me apresentaram o PPGBEA e a beleza da estatística e ao professor Ricardo de Carvalho, exemplo de profissional, fonte de inspiração profissional, por todo conhecimento e pelo apoio durante a graduação.

Aos professores, diretor do departamento, coordenador e secretário PPGBEA por todo conhecimento, apoio e disponibilidade. A todos os funcionários do departamento que viabilizam espaço, instrumentos e limpeza para que nós alunos pudéssemos desenvolver nossas pesquisas.

Agradeço ao meu orientador, professor Dr. Frank Sinatra G. da Silva, por todo conhecimento, paciência, disponibilidade e apoio.

Agradeço ao meu coorientador, professor Dr. Paulo Renato A. Firmino, inicialmente pela oportunidade de ser membro do grupo de pesquisa Modelagem Estatística, Simulação e Otimização de Risco (MESOR) e por todo conhecimento, paciência, disponibilidade, apoio

e incentivo. Aos colegas do grupo MESOR, Jair e Eduardo, pelas dúvidas solucionadas e parceria nos estudos.

A CAPES pelo apoio financeiro.

A Seduc-CE pela concessão da licença para estudos. Aos colegas professores, gestão e secretaria da E.E.E.P. Prof. José Osmar Plácido da Silva que, muito solícitos, me ajudaram no início do mestrado.

Enfim, a todos com quem tive a honra de conviver e que, de forma direta ou indireta, contribuíram para a construção desse trabalho.

*“Imagine all the people  
Living life in peace  
You may say, I’m a dreamer  
But I’m not the only one  
I hope someday you’ll join us  
And the world will be as one”  
(Imagine – John Lennon)*

# Resumo

O século XXI é marcado por discussões acerca da promoção do desenvolvimento sustentável e sua importância no contexto mundial. Para tal, mostra-se indispensável o uso de indicadores de controle, que permitem monitorar, entender, diagnosticar e prever variáveis de interesse. Dessa forma, a modelagem e previsão de séries temporais são uma importante ferramenta de tomada de decisão para o planejamento e adequada execução de ações nos três eixos elementares do desenvolvimento sustentável: ambiental, econômico e social. A modelagem estatístico-computacional tem se tornado cada dia mais eficiente com o desenvolvimento de novas técnicas para analisar o comportamento de fenômenos e suas mudanças ao longo do tempo, bem como prever seus valores futuros com base em observações passadas. E, entre essas técnicas de previsão, destaca-se as combinações de preditores, que têm se mostrado estatisticamente mais acuradas e eficientes que modelos individuais, por exemplo. Abordagens como a combinação de preditores via cópulas e SVR (sigla para regressão de vetor de suporte, do inglês *support vector regression*) são promissoras nessa área. Há ainda as abordagens mais clássicas, como a combinação por média simples, mediana simples ou mínima variância. Este trabalho tem como objetivo realizar um estudo do desempenho destes combinadores na previsão de séries temporais úteis ao desenvolvimento sustentável. Ao todo, doze séries temporais são estudadas, a partir de quatro formalismos de modelagem individual e quatro de combinação. Em todos os casos, modelos otimizados são desenvolvidos, a partir de algoritmos de otimização. O percurso metodológico consiste basicamente em: i) construir modelos individuais para cada série baseados em formalismos como redes neurais artificiais, modelos auto regressivos e de médias móveis, de alisamento exponencial e regressão de vetores de suporte; ii) construir os combinadores e iii) avaliar o desempenho dos modelos, a partir de um conjunto de medidas de qualidade consagradas na literatura, bem como outras sugeridas neste trabalho. Os resultados mostram que os combinadores via cópula são mais promissores para séries úteis ao desenvolvimento sustentável quando comparados aos combinadores por SVR. Ainda, mesmo diante de modelos combinados os modelos mais simples apresentaram bons resultados. A parcimônia e o superajuste ao treinamento podem ter comprometido o desempenho dos modelos de aprendizado de máquina individuais, como SVR e ANN, afetando assim o desempenho dos combinadores.

**Palavras-chave:** Modelagem e previsão de séries temporais. Combinação de preditores. Desenvolvimento sustentável. Cópulas. Regressão de vetor de suporte.

# Abstract

The 21st century has emphasized discussions regarding the sustainable development. In this way, it has been paramount the use of control indicators aimed to monitor, understand, diagnostic, and forecast variables of interest. Thus, time series modelling and forecasting exercises are important tools for decision making with respect to plans and actions regarding the triple bottom line of the sustainable development: environment, economy, and society. The statistical-computing modelling has become attractive by means of the development of technics for analyzing and predicting the behavior of phenomena changing through the time, considering the respective past observations. Among the alternatives, one must highlight the combination of predictors. Usually, they have been more accurate and efficient in statistical terms in comparison with single models. Therefore, approaches like copulas and support vector regression (SVR) seem useful, though classical formalisms, such as simple average, median, and minimal variance combinator, have presented good results. This work aims to compare the performance of the aforementioned combination strategies when forecasting time series useful for the sustainable development. Thus, twelve time series are taken into account, as well as four single modeling and four combination modeling formalisms. For all cases, optimized models are provided via optimization methods. The methodology is based on: i) to build single models for each time series, considering artificial neural nets, autoregressive and moving average, exponential smoothing, and SVR; ii) to build the combinator; and iii) to evaluate the performance of the models, according to a number of established quality metrics as well as measures proposed in the work. The results show that copula combinator are more promising for series useful for sustainable development when compared to SVR combinator. Still, even in the face of combined models, the simplest models showed good results. Parsimony and overfitting training may have compromised the performance of individual machine learning models, such as SVR and ANN, thus affecting the performance of the combiners.

**Key-words:** Time series modelling and forecasting. Predictors combination. Sustainable development. Copulas. Support vector regression.

# Lista de Figuras

Figura 1 – Representação básica do Neurônio Biológico . . . . .	12
Figura 2 – Modelagem básica do Neurônio Artificial . . . . .	13
Figura 3 – Arquitetura de Rede Neural adaptada para séries temporais com duas entradas regressas. . . . .	14
Figura 4 – Funções de Ativação Linear (a); Sigmóide Logística (b) e Tangente Hiperbólica (c) . . . . .	16
Figura 5 – Função Perda $\epsilon$ -insensível. . . . .	20
Figura 6 – SVR para séries temporais. . . . .	21
Figura 7 – Arquitetura do processo de treinamento para SVR. . . . .	23
Figura 8 – Arquitetura da Combinação via Cópula. . . . .	32
Figura 9 – Arquitetura da Combinação via SVR. . . . .	33
Figura 10 – Arquitetura proposta para modelagem individual e combinada. . . . .	40
Figura 11 – Séries Temporais Ambientais. . . . .	43
Figura 12 – Séries Temporais Econômicas. . . . .	44
Figura 13 – Séries Temporais Sociais. . . . .	45
Figura 14 – Média Agregada de Performance dos modelos individuais e combinados para a série DEF. . . . .	49
Figura 15 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série DEF (Fase de teste). . . . .	49
Figura 16 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série DEF (Fase de teste). . . . .	50
Figura 17 – Média Agregada de Performance dos modelos individuais e combinados para a série ONI. . . . .	52
Figura 18 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série ONI (Fase de teste). . . . .	52
Figura 19 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série ONI (Fase de teste). . . . .	53
Figura 20 – Média Agregada de Performance dos modelos individuais e combinados para a série CO <sub>2</sub> . . . . .	55
Figura 21 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série CO <sub>2</sub> (Fase de teste). . . . .	56

Figura 22 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série CO <sub>2</sub> (Fase de teste). . . . .	56
Figura 23 – Média Agregada de Performance dos modelos individuais e combinados para a série SUN. . . . .	58
Figura 24 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série SUN (Fase de teste). . . . .	59
Figura 25 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série SUN (Fase de teste). . . . .	59
Figura 26 – Média Agregada de Performance dos modelos individuais e combinados para a série BBP. . . . .	61
Figura 27 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série BBP (Fase de teste). . . . .	62
Figura 28 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série BBP (Fase de teste). . . . .	62
Figura 29 – Média Agregada de Performance dos modelos individuais e combinados para a série IPCA. . . . .	64
Figura 30 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série IPCA (Fase de teste). . . . .	65
Figura 31 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série IPCA (Fase de teste). . . . .	65
Figura 32 – Média Agregada de Performance dos modelos individuais e combinados para a série PSP. . . . .	67
Figura 33 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série PSP (Fase de teste). . . . .	68
Figura 34 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série PSP (Fase de teste). . . . .	68
Figura 35 – Média Agregada de Performance dos modelos individuais e combinados para a série SP500. . . . .	70
Figura 36 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série SP500 (Fase de teste). . . . .	71
Figura 37 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série SP500 (Fase de teste). . . . .	71
Figura 38 – Média Agregada de Performance dos modelos individuais e combinados para a série MVF. . . . .	73
Figura 39 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série MVF (Fase de teste). . . . .	74

Figura 40 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série MVF (Fase de teste). . . . .	74
Figura 41 – Média Agregada de Performance dos modelos individuais e combinados para a série DEN. . . . .	76
Figura 42 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série DEN (Fase de teste). . . . .	77
Figura 43 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série DEN (Fase de teste). . . . .	77
Figura 44 – Média Agregada de Performance dos modelos individuais e combinados para a série ENE. . . . .	79
Figura 45 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série ENE (Fase de teste). . . . .	80
Figura 46 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série ENE (Fase de teste). . . . .	80
Figura 47 – Média Agregada de Performance dos modelos individuais e combinados para a série MUC. . . . .	82
Figura 48 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série MUC (Fase de teste). . . . .	82
Figura 49 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série MUC (Fase de teste). . . . .	83

## Lista de tabelas

Tabela 1 – Classificação ETS. . . . .	11
Tabela 2 – Resumo das Características das Séries Temporais em Estudo . . . . .	42
Tabela 3 – Parâmetros de ajuste dos modelos. . . . .	46
Tabela 4 – Descrição dos modelos individuais e combinados para a série DEF. . . . .	48
Tabela 5 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série DEF (fase de teste). Os melhores valores encontram-se em negrito. . . . .	48
Tabela 6 – Descrição dos modelos individuais e combinados para a série ONI. . . . .	51
Tabela 7 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série ONI (fase de teste). Os melhores valores encontram-se em negrito. . . . .	51
Tabela 8 – Descrição dos modelos individuais e combinados para a série CO <sub>2</sub> . . . . .	54
Tabela 9 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série CO <sub>2</sub> (fase de teste). Os melhores valores encontram-se em negrito . . . . .	55
Tabela 10 – Descrição dos modelos individuais e combinados para a série SUN. . . . .	57
Tabela 11 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série SUN (fase de teste). Os melhores valores encontram-se em negrito. . . . .	58
Tabela 12 – Descrição dos modelos individuais e combinados para a série BBP. . . . .	60
Tabela 13 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série BBP (fase de teste). Os melhores valores encontram-se em negrito. . . . .	61
Tabela 14 – Descrição dos modelos individuais e combinados para a série IPCA. . . . .	63
Tabela 15 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série IPCA (fase de teste). Os melhores valores encontram-se em negrito. . . . .	64
Tabela 16 – Descrição dos modelos individuais e combinados para a série PSP. . . . .	66
Tabela 17 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série PSP (fase de teste). Os melhores valores encontram-se em negrito. . . . .	67
Tabela 18 – Descrição dos modelos individuais e combinados para a série SP500. . . . .	69

Tabela 19 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série SP500 (fase de teste). Os melhores valores encontram-se em negrito. . . . .	70
Tabela 20 – Descrição dos modelos individuais e combinados para a série MVF. . . . .	72
Tabela 21 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série MVF (fase de teste). Os melhores valores encontram-se em negrito. . . . .	73
Tabela 22 – Descrição dos modelos individuais e combinados para a série DEN. . . . .	75
Tabela 23 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série DEN (fase de teste). Os melhores valores encontram-se em negrito. . . . .	76
Tabela 24 – Descrição dos modelos individuais e combinados para a série ENE. . . . .	78
Tabela 25 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série ENE (fase de teste). Os melhores valores encontram-se em negrito. . . . .	79
Tabela 26 – Descrição dos modelos individuais e combinados para a série MUC. . . . .	81
Tabela 27 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série MUC (fase de teste). Os melhores valores encontram-se em negrito. . . . .	81
Tabela 28 – Desempenho normalizado médio agregado dos modelos (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) ao prever as séries DEF, CO <sub>2</sub> , SUN, ONI, BBP, SP500, IPCA, PSP, MVF, MUC, ENE, DEN (Fase de teste). Os melhores valores encontram-se em negrito. . . . .	84
Tabela 29 – Desempenho normalizado médio agregado dos modelos (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) ao prever as séries DEF, CO <sub>2</sub> , SUN, ONI, BBP, SP500, IPCA, PSP, MVF, MUC, ENE, DEN (Fase de treinamento). Os melhores valores encontram-se em negrito. . . . .	86
Tabela 30 – Tempo de treinamento dos modelos (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para as séries DEF, CO <sub>2</sub> , SUN, ONI, BBP, SP500, IPCA, PSP, MVF, MUC, ENE, DEN (em segundos). Os melhores valores encontram-se em negrito. . . . .	87

# Lista de abreviaturas e siglas

AF	Função de Ativação
AM	Aprendizado de Máquina
ANN	Redes Neurais Artificiais
AR	Auto-regressivo
ARIMA	Auto-regressivo Integrado de Médias Móveis
ARMA	Auto-regressivo de Médias Móveis
ARV	Variância Relativa a Média
BBP	Produção Brasileira de Biodiesel
BIC	Critério de Informação Baysiano
BPROP	Retroprogramação
CA	Cópulas Arquimedianas
cCB	Combinação via Cópula
CE	Cópulas Elípticas
CO <sub>2</sub>	Gás Carbônico
CO2	Concentração de Dióxido de Carbono na Atmosfera
cSA	Combinação por Média Simples
cSM	Combinação por Mediana Simples
cSVM	Combinação via <i>Support Vector Machines</i>
DEF	Desmatamento da Amazônia Brasileira
DEN	Número de Casos de Dengue no Rio de Janeiro
DS	Desenvolvimento Sustentável

EMV	Estimador de Máxima Verossimilhança
ENE	Consumo de Energia no Nordeste Brasileiro
ENSO	El Niño-Oscilação do Sul
ETS	Suavização Exponencial
GA	Algoritmo Genético
GRPROP	Algoritmo Globalmente Convergente
IBGE	Instituto Brasileiro de Geografia e Estatística
IC	Índice de Concordância
ID	Índice de Desacordo
IDH	Índice de Desenvolvimento Humano
IFM	<i>Inference for Marginals</i> ou Inferência por Marginais
IPCA	Índice de Preços ao Consumidor no Brasil
LA	Algoritmo de aprendizagem
LF	Função Logística
LK	Função Kernel Linear
MA	Médias Móveis
MAN	Média Agregada Normalizada
MAPE	Erro Médio Absoluto
MLP	Rede Multilayer Perceptron
MMV	Método da Máxima Verossimilhança
MSE	Erro Quadrático Médio
MUC	Número de Casos de Caxumba em Nova York
MV	Mínima Variância
MVF	Mortes em Veículos Motorizados
ODS	Objetivos de Desenvolvimento Sustentável

ONI	Índice Oceânico de Niño
ONU	Organização das Nações Unidas
PIB	Produto Interno Bruto
PK	Função Kernel Polinomial
PSO	Otimização por Enxame de Partículas
PSP	Preço das Ações da Petrobras
RBFK	Função Kernel de Base Radial
RPROP	Retropropagação Resiliente
SA	<i>Simulated Annealing</i> ou Recozimento Simulado
SARIMA	Auto-regressivo Integrado de Médias Móveis Sazonal
SK	Função Kernel Sigmóide
SP500	Índice SP500
SUN	Índice de Manchas Solares
SV	Vetores de Suporte
SVM	<i>Support Vector Machines</i> ou Máquina de Vetores de Suporte
SVR	<i>Support Vector Regression</i> ou Regressão de Vetores de Suporte
Theil'U	Coefficiente U de Theil
THF	Função Tangente Hiperbólica
WPOCID	Predição de Alteração na Direção
WR <sup>2</sup>	Coefficiente de Indeterminação

# Lista de Símbolos

$\alpha_i, \alpha_i^*$	Multiplicadores de Lagrange
$\Delta E$	Variação da função objetivo (SA)
$\Delta$	Operador diferença (ARIMA)
$\epsilon$	Desvio
$\eta^+, \eta^-$	Fatores de atualização de pesos LA
$\gamma$	Vetor de parâmetros da cópula
$\lambda$	Parâmetro da KF
$\mathbf{e}$	Vetor de resíduos
$\mathbf{E}_t$	Vetor de erros dos preditores individuais
$\mathbf{T}$	Produto interno
$\mathbf{w}$	Vetor de pesos
$\mathbf{X}$	Vetor de preditores individuais
$\mathbf{x}_t$	Vetor de estimativas correspondentes aos preditores individuais
$\mathcal{A}$	Axônio
$\mathcal{A}_h(\cdot)$	Função de Ativação
$\mathcal{I}$	Dentritos
$\mathcal{N}$	Neurônio artificial
$\mathcal{P}$	Perceptron
$\mathcal{S}$	Corpo celular
$\mathcal{S}_h(\cdot)$	Função de Ativação Intermediária
$\mu$	Valor constante para a Média

$\phi_i$	Coeficientes auto-regressivos (AR)
$\Phi(\cdot)$	Operador auto-regressivo (ARIMA); Distribuição Normal Padrão Multivariada (cCB)
$\rho$	Matriz de correlação
$\sigma^2$	Valor constante para a Variância
$\Theta$	Operador de médias móveis
$\theta_i$	Coeficientes de médias móveis (MA); Coeficiente que pondera as observações $u_{tt_i}$ (ANN); Parâmetro de dependência (cCB)
$\Upsilon$	Valor constante para a Covariância
$\hat{u}_t$	Valor estimado para $u_t$ no instante $t$
$\xi_i, \xi_i^*$	Variáveis de folga
$A$	Aditivo/a
$Ad$	Aditiva amortecida
$AR(p)$	Modelo Auto-regressivo
$ARIMA(p, d, q)$	Modelo Auto-regressivo Integrado de Médias Moveis
$ARIMA(p, d, q)(P, D, Q)$	Modelo Auto-regressivo Integrado de Médias Móveis Sazonal
$ARMA(p, q)$	Modelo Auto-regressivo de Médias Móveis
$B$	Operador translação para o passado
$b$	Viés
$C$	Constante de regularização
$C(\cdot)$	Fução Cópula
$c(\cdot)$	Fução Densidade de Probabilidade da cópula
$coef0$	Intercepto das Funções PK e SK
$Cov(\cdot)$	Fução de Covariância
$d$	Ordem de diferenciação

<i>degree</i>	Grau do polinômio (PK)
$E$	Erro modelo ETS
$E(\cdot)$	Função Média
$E_t$	Erro aleatório
$E_{t,i}$	Entrada do vetor $\mathbf{E}$
$e_{t,i}$	Entrada do vetor $\mathbf{e}$
$e_{t,i}$	Resíduo
$ETS(E, T, S)$	Modelo de Suavização Exponencial
$f(\mathbf{x})$	Valor predito pela função de aproximação (SVM)
$F^{-1}(\cdot)$	Função de Distribuição Acumulada Marginal Inversa
$H$	Quantidade de neurônios
$h$	Índice do neurônio
$H(\cdot)$	Função de Distribuição Acumulada Conjunta
$K(\cdot)$	Função Kernel
$L$	Função de Verossimilhança
$l$	Função de Log-verossimilhança
$L(\cdot)$	Função de Perda $\epsilon$ -insensível
$M$	Multiplicativo/a
$MA(q)$	Modelo de Médias Móveis
$Md$	Multiplicativa amortecida
$N$	Nenhuma
$n$	Quantidade de observações da amostra
$NHL$	Quantidade de nodos na camada oculta (ANN)
$NI$	Número de entradas auto-regressivas (ANN)
$NO$	Número de saídas (ANN)

$p$	Ordem do modelo auto-regressivo (ARIMA); Solução corrente (SA)
$p(\cdot)$	Função Densidade de Probabilidade
$p_0$	Solução inicial
$P_{ARS}$	Número de parâmetros auto-regressivos sazonais
$P_{AR}$	Número de parâmetros auto-regressivos
$p_{min}$	Solução ótima
$q$	Ordem do modelo de médias móveis
$S$	Sazonalidade
$T$	Tendência (ETS); Temperatura (SA)
$t$	Índice cronológico
$T_0$	Temperatura inicial
$U_t$	Processo estocástico
$u_t$	Valor observado da série temporal no instante $t$
$u_t^{SA}$	Valor predito pela combinação cSA
$V(\cdot)$	função Variância
$v_i, F(\cdot)$	Função de Distribuição Acumulada Marginal
$X$	Vetor aleatório
$x_{i,t}$	Entrada do vetor $\mathbf{x}_t$
$X_{t,i}$	Entrada do vetor $\mathbf{X}$
$\theta_0$	Intercepto ANN
$\varphi(\cdot)$	Função Geradora da CA

# List of Algorithms

1	Pseudocódigo Algoritmo Backpropagation . . . . .	17
2	Pseudocódigo Algoritmo <i>Simulated Annealing</i> adaptado. . . . .	34

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Objetivos	3
1.1.1	Geral	3
1.1.2	Específicos	3
1.2	Estrutura da dissertação	4
<b>2</b>	<b>Revisão de Literatura</b>	<b>5</b>
2.1	Desenvolvimento Sustentável	5
2.2	Séries Temporais	6
2.3	Modelos Box e Jenkins	8
2.3.1	Modelos Auto-regressivos de Ordem $p$	8
2.3.2	Modelos de Médias Móveis de Ordem $q$	8
2.3.3	Modelos Auto-regressivos de Médias Móveis	9
2.3.4	Modelos Auto-regressivos Integrados de Médias Móveis	9
2.4	Modelos de Suavização Exponencial	10
2.5	Redes Neurais Artificiais	11
2.5.1	Rede Multilayer Perceptron – MLP	12
2.5.2	Funções de Ativação	15
2.5.3	Algoritmos de Aprendizado	16
2.6	Regressão de Vetor de Suporte	18
2.6.1	Funções Kernel	21
2.7	Formalismos de Modelagem Combinada	24
2.7.1	Média Simples e Mediana Simples	25
2.7.2	Cópulas	25
2.7.3	Combinação via Cópulas	30
2.7.4	Combinação via <i>Support Vector Regression</i>	33
2.8	Algoritmo <i>Simulated Annealing</i>	33
2.9	Métricas de Desempenho	36
2.10	Crítério de Informação Bayesiano	38
<b>3</b>	<b>Experimentação</b>	<b>40</b>
3.1	Arquitetura do estudo	40
3.2	Conjunto de Dados	42
<b>4</b>	<b>Resultados</b>	<b>46</b>

4.1	Séries Temporais Ambientais . . . . .	47
4.1.1	Desmatamento da Amazônia Brasileira . . . . .	47
4.1.2	Índice Oceânico de <i>Niño</i> . . . . .	50
4.1.3	Concentração de Dióxido de Carbono na Atmosfera . . . . .	53
4.1.4	Índice de Manchas Solares . . . . .	56
4.2	Séries Temporais Econômicas . . . . .	60
4.2.1	Produção Brasileira de Biodiesel . . . . .	60
4.2.2	Índice de Preços ao Consumidor no Brasil . . . . .	62
4.2.3	Preço de ações da Petrobras . . . . .	65
4.2.4	Índice <i>Standard &amp; Poor's 500</i> . . . . .	68
4.3	Séries Temporais Sociais . . . . .	72
4.3.1	Mortes por Veículos Motorizados . . . . .	72
4.3.2	Número de Casos de Dengue no Rio de Janeiro . . . . .	75
4.3.3	Consumo de Energia no Nordeste Brasileiro . . . . .	77
4.3.4	Número de Casos de Caxumba em Nova York . . . . .	80
5	Discussão . . . . .	84
6	Conclusão . . . . .	88
6.1	Limitações da Pesquisa . . . . .	88
6.2	Trabalhos Futuros . . . . .	89
	Referências Bibliográficas . . . . .	90

# 1 Introdução

Cada vez mais, os termos responsabilidade social, responsabilidade corporativa, sustentabilidade e governança se entrelaçam e não raramente são tratados como sinônimos (CROWTHER; SEIFI; MOYEEN, 2018). Estes conceitos estão intimamente ligados ao tema Desenvolvimento Sustentável (DS), termo formalizado pela primeira vez no Relatório de Brundtland, da Comissão Mundial sobre Meio Ambiente e Desenvolvimento criada em 1983 pela Organização das Nações Unidas (ONU), que documentou cuidadosamente os estados atual e futuro econômico e ecológico em esfera global (MITCHAM, 1995).

Segundo o Relatório de Brundtland, Desenvolvimento Sustentável, em sua definição mais usual, é aquele que busca satisfazer as necessidades do presente, sem comprometer a capacidade das gerações futuras de satisfazerem as suas próprias necessidades. Com isso, o DS implica em limites, limitações aos recursos ambientais e a capacidade da biosfera de absorver os impactos das atividades humanas (MITCHAM, 1995).

Esse conceito surgiu na busca de solucionar o conflito entre as metas ambientais e o desenvolvimento socio-econômico. Desde então, houve o reconhecimento de três aspectos do Desenvolvimento Sustentável: Ambiental, Econômico e Social, dimensões que sugerem de maneira clara novas diretrizes para o processo de desenvolvimento, no qual o DS ganha um novo prisma: deve reparar as desigualdades sociais e os danos ambientais, mantendo uma base econômica sólida (HARRIS, 2001).

As discussões acerca do DS apontaram para a necessidade de indicadores confiáveis e pertinentes para orientar o seu processo, uma vez que os indicadores existentes não forneciam designações adequadas de sustentabilidade, estes devem ser um instrumento de avaliação democrática e uma ferramenta de gestão governamental (BOULANGER, 2008).

De maneira simples, Box, Jenkins e Reinsel (2008) define uma série temporal como uma sequência de observações ordenadas no tempo, e sua modelagem e previsão se apresentam como uma importante ferramenta na tomada de decisões em diversos campos; não obstante, pode ser útil no estudo de variáveis ligadas ao DS, uma vez que possibilita a elaboração de modelos que descrevam adequadamente o comportamento destas através do tempo, objetivando identificar mudanças ou prever seu comportamento futuro. Trabalhos como os de Joshi, Hughes e Sisk (2015) e Boontome, Therdyothin e Chontanawat (2019) corroboram essa ideia.

No contexto de séries temporais, têm-se empregado uma gama de metodologias, as mais comuns são os modelos lineares como o Autorregressivo Integrado de Médias Móveis (ARIMA) [Box, Jenkins e Reinsel \(2008\)](#) e Suavização Exponencial (ETS) [Hyndman e Koehler \(2008\)](#) aqui abordadas. Outra proposta recorrente no campo de *Machine Learning*, do português Aprendizado de Máquina, são Redes Neurais Artificiais (ANN) [Haykin \(2001\)](#) [Braga, Carvalho e Ludemir \(2007\)](#) e *Support Vector Machines* (SVM) [Vapnik \(2000\)](#), que têm mostrado resultados positivos.

Na literatura, é comum procurar um modelo único inerente à série. Porém, este modelo pode não ser encontrado e, em sua busca, pode-se subestimar seriamente a incerteza nas previsões ([CHATFIELD, 2000](#)). Entretanto, tem sido frequente a busca por novas técnicas de predição, agora baseadas na combinação de preditores. Estes formalismos combinam modelos individuais diversos, de forma a promover melhores previsões da série em questão. Geralmente, a agregação é feita por uma combinação linear, a exemplo das alternativas Média Simples e Mediana Simples e Mínima Variância. Sob este prisma, mesmo indiretamente, supõe-se normalidade entre os erros dos modelos individuais. Segundo [Firmino, Neto e Ferreira \(2014\)](#), os preditores combinados têm estado na vanguarda da modelagem de séries temporais.

A combinação de modelos de previsão para séries temporais tem se mostrado promissora pelo fato de usualmente promover estimadores estatisticamente mais acurados e eficientes que os modelos individuais ([OLIVEIRA, 2017](#)). Abordagens como a combinação de preditores via cópulas e SVM são atraentes.

Pesquisadores como [Assis \(2017\)](#) e [Oliveira \(2014\)](#) trazem importante contribuição nessa área quando fazem uso de cópulas. Nessa abordagem, os erros podem ou não ser normalmente distribuídos, uma vez que a modelagem via cópula se utiliza das distribuições marginais para formar a distribuição conjunta e permite conhecer a relação de dependência entre os preditores combinados. Assim, os combinadores via cópulas podem ser não-lineares. Estudos como os de [Firmino \(2016\)](#) e [Oliveira \(2018\)](#) mostram que a combinação de modelos via cópulas apresenta em alguns casos resultados superiores quando comparados aos modelos individuais e aos modelos agregados por combinação linear, a saber, métodos de média simples (SA) e de mínima variância (MV).

Cópulas são funções que combinam distribuições univariadas para construir uma distribuição multivariada. A teoria associada ao estudo de cópulas é baseada nos trabalhos de Sklar [Sklar \(1959\)](#) à sombra do teorema que leva seu nome. O interesse estatístico em cópulas advém da capacidade que possuem de incorporar a dependência entre as distribuições univariadas e construir novas famílias de distribuições multivariadas ([NELSEN, 2013](#)).

Alternativamente, tem-se empregado SVM em sua forma adaptada para regressão *Support Vector Regression* (SVR) [Smola e Scholkopf \(2004\)](#), na previsão de séries temporais. Proposto por [Vapnik \(2000\)](#), o algoritmo tinha como objetivo inicial a classificação de dados categóricos. Trabalhos como os de [Bozic e Stojanovic \(2011\)](#), para demanda de eletricidade e [Kim \(2003\)](#) e [Okasha \(2014\)](#), para dados financeiros, demonstram uma boa qualidade do método, por superar problemas como a não-linearidade e a não-estacionaridade, além de oferecer vantagens importantes sobre outros métodos, como ter um número menor de parâmetros e produzir previsões mais precisas ([OKASHA, 2014](#)).

Nesta dissertação é proposta a modelagem e previsão de séries temporais ligadas ao desenvolvimento sustentável através de formalismos individuais e combinados. Será apresentada uma estrutura metodológica alternativa para obter previsões a partir de modelos individuais e combinados otimizados, com especial atenção a SVM e cópulas. As combinações via SVM e cópula serão comparadas com SA e SM, a partir de um conjunto de medidas de desempenho.

Com base nesse encadeamento de ideias, buscou-se responder ao questionamento: Os combinadores baseados em SVM apresentam melhores previsões, inclusive em comparação com aqueles baseados em cópulas?

## 1.1 Objetivos

### 1.1.1 Geral

Comparar o desempenho de SVM, cópulas e outros formalismos, em exercícios envolvendo a combinação de preditores de séries temporais.

### 1.1.2 Específicos

- (i) Estudar alternativas consagradas de modelagem de séries temporais, tais como modelos autoregressivos integrado e de média móvel (ARIMA), redes neurais artificiais (ANN) e alisamento exponencial (ETS), bem como pacotes computacionais correlatos.
- (ii) Estudar a teoria de SVM e cópulas, bem como pacotes computacionais correlatos.
- (iii) Implementar as alternativas de modelagem de (i) e combinações a partir de (ii), para uso em escala.
- (iv) Analisar comparativamente os resultados obtidos pelos preditores, a partir de casos do mundo real.

## 1.2 Estrutura da dissertação

Além deste capítulo, esta dissertação envolve mais cinco:

**Capítulo 2 - Revisão de Literatura:** apresenta um aporte teórico acerca do desenvolvimento sustentável, de séries temporais, dos formalismos de modelagem individuais e combinados e medidas de desempenho dos modelos estudados.

**Capítulo 3 - Experimentação:** apresenta a arquitetura para o estudo realizado e os dados reais para os estudos de casos acompanhados de suas descrições.

**Capítulo 4 - Resultados:** apresenta os resultados da modelagem individual e combinada para as séries temporais trabalhadas.

**Capítulo 5 - Discussão:** apresenta as discussões necessárias de forma geral.

**Capítulo 6 - Conclusões:** apresenta as conclusões do estudo realizado, limitações do trabalho e a proposta para futuros trabalhos.

## 2 Revisão de Literatura

### 2.1 Desenvolvimento Sustentável

O termo Desenvolvimento Sustentável foi citado pela primeira vez em 1987 no Relatório de Brundtland, intitulado *Nosso Futuro Comum*, apresentado pela Comissão Mundial de Meio Ambiente e Desenvolvimento. Em seu texto, o DS é definido como “aquele que atende às necessidades do presente sem comprometer a possibilidade de que as gerações futuras atendam as suas próprias necessidades”. Esse relatório examina questões críticas ligadas ao meio ambiente e ao desenvolvimento apresentando sugestões plausíveis para lidar com estes temas em nível mundial (KEEBLE, 1988).

O conceito de DS surge como um tema ambiental importante e um forte paradigma: se por um lado DS significa desenvolvimento econômico aliado a um padrão de vida que não comprometa o meio ambiente para que as próximas gerações tenham recursos suficientes para viver, por outro lado, é difícil enxergar meios de relacioná-lo com o estilo de vida da população. De fato, ninguém sabe ao certo o que é o DS sob a premissa que não se pode citar exemplos onde este ocorreu; com efeito, este deve incluir uma grande transformação da sociedade (MUSCHETT, 1996).

Nas discussões acerca da temática DS houve o reconhecimento de três aspectos chave: o econômico, o ambiental e o social, de forma que a realização do DS não é possível sem levar em consideração essas três questões concomitantemente. De fato, o DS trata-se de um modelo de desenvolvimento comprometido em harmonizar ações de curto, médio e longo prazos, sempre entrelaçando essas três dimensões (HARRIS, 2001; VESCO; FERRERO, 2015).

Na dimensão econômica, idealiza-se um sistema capaz de produzir bens e serviços de forma contínua, acessível e ambientalmente responsável. No sentido ambiental, busca-se o uso inteligente de recursos naturais, equilibrando o seu consumo e sua manutenção, na geração de oportunidades de negócio e de preservação. No aspecto social do DS, deve-se considerar a equidade na distribuição de renda, adequada prestação de serviços sociais, igualdade de gêneros e promoção de políticas sociais (HARRIS, 2001).

Em 2015, as Nações Unidas, com o intuito de apoiar a transição para um desenvolvimento sustentável, definiram metas para governos e empresas chamadas Objetivos de Desenvolvimento Sustentável (ODS). São 17 objetivos que ficaram conhecidos como a

Agenda 2030 ou Agenda Mundial para o DS e fornecem uma estrutura coerente, holística e integrada para trilhar o caminho até a sustentabilidade de forma a enfrentar os desafios mais urgentes criando um futuro melhor (GRI, 2015).

O ideal de estreitar as relações e colaborações entre negócios, política, sociedade e ambiente natural não é mais opcional e passa a ser visto como uma necessidade no âmbito do Desenvolvimento Sustentável. As Nações foram convidadas a colaborar, planejar e implementar um caminho de encontro ao DS, preocupando-se com as mudanças e a promoção do progresso econômico e social em seus Estados (CROWTHER; SEIFI; MOYEEN, 2018).

No contexto de avaliar a sustentabilidade dos sistemas, surgiu a necessidade de criar novos indicadores direcionados para o DS, uma vez que os existentes não ofereciam informações adequadas nesse aspecto e, portanto não poderiam servir como base para tomada de decisões no campo do DS. Com o tempo, esses indicadores passaram a ser tratados como Indicadores Sociais, definidos por Bauer (1966) como estatísticas, séries estatísticas e todas as outras formas de evidência que permitem avaliar o estado atual e onde se quer chegar em relação aos valores e objetivos. Como exemplo desses indicadores tem-se o Índice de Desenvolvimento Humano (IDH) (BOULANGER, 2008).

Nos últimos anos várias pontes entre os ODS e a ciência têm sido construídas, despertando uma mobilização de forma técnica e científica da academia, sociedade e setor privado para apoiar as causas do DS e intervirem com soluções em escalas local, nacional e global (ASSEMBLY, 2015).

Nesse cenário, a abordagem de séries temporais tem sido aplicada por diversos pesquisadores em situações ligadas ao DS. São exemplos os estudos de Wu et al. (2019) que usou modelos de redes neurais para avaliar se o desenvolvimento urbano é sustentável e Magazzino (2016) que usaram uma abordagem de séries temporais para verificar a relação entre Produto Interno Bruto (PIB) real, emissões de CO<sub>2</sub> e uso de energia.

## 2.2 Séries Temporais

Uma série temporal é um conjunto de observações ordenadas no tempo. Na prática, uma série temporal é construída a partir de observações e/ou medidas de um certo fenômeno. São exemplos: quantidade de chuvas mensais, número de acidentes rodoviários semanais, observações horárias de um processo químico, entre outros (BOX; JENKINS; REINSEL, 2008).

Uma série temporal pode ser definida como

$$\{u_t \in \mathbb{R} | t = 1, 2, 3, \dots, n\}, n \in \mathbb{N}, \quad (2.1)$$

no qual  $t$  denota o índice cronológico e  $n$  a quantidade de observações da amostra.

Um Modelo Estocástico ou simplesmente Processo Estocástico é um modelo que descreve a estrutura de probabilidades de uma sequência de observações. Dentre esses modelos, uma classe importante são os Processos Estacionários, que se encontram em um estado de equilíbrio estatístico. Uma série temporal é uma amostra de um processo estocástico indexado no tempo e uma maneira de descrevê-la seria determinar todos os seus momentos; porém, o interesse são os momentos de baixa ordem (BOX; JENKINS; REINSEL, 2008).

Uma série temporal é dita estritamente estacionária se suas propriedades não são afetadas por uma mudança na origem do tempo. Em outras palavras, quando a distribuição de probabilidades conjunta das variáveis subjacentes às observações  $u_t, u_{t+1}, \dots, u_{t+n}$  é exatamente igual à distribuição de probabilidades conjunta daquelas que regem as observações  $u_{t+k}, u_{t+k+1}, \dots, u_{t+k+n}$  (MONTGOMERY; JENNINGS; KULAHCI, 2015).

Um processo estocástico é estacionário de segunda ordem se as condições abaixo forem satisfeitas para todo instante de tempo  $t$  (BOX; JENKINS; REINSEL, 2008):

$$E[U_t] = E[U_{t+k}] = \mu, \quad (2.2)$$

$$V[U_t] = V[(U_t - \mu)^2] = \sigma^2, \quad (2.3)$$

$$Cov[U_t, U_{t+k}] = E[(U_t - \mu) \cdot (U_{t+k} - \mu)] = \Upsilon, \quad (2.4)$$

no qual  $E[U_t]$  é a função média,  $V[U_t]$  é a função variância e  $Cov[U_t, U_{t+k}]$  é a função de covariância e  $\mu$ ,  $\sigma$  e  $\Upsilon$  são constantes no tempo.

O objetivo da previsão de séries temporais é estimar um valor futuro baseado em observações regressas da série, de forma que o valor predito seja imparcial e consistente. Os estimadores podem ser divididos em dois tipos: os lineares e os não lineares. (SAPANKEVYCH; SANKAR, 2009).

Como exemplos de métodos lineares são apresentados o auto-regressivo Integrado de Médias Móveis e os Modelos de Suavização Exponencial clássicos. Já outras técnicas de *Machine Learning*, como Redes Neurais Artificiais e *Support Vector Machines* são do tipo não lineares.

## 2.3 Modelos Box e Jenkins

Uma metodologia bastante utilizada na modelagem de séries temporais é a abordagem [Box, Jenkins e Reinsel \(2008\)](#). Esse método consiste em ajustar modelos Auto-regressivos Integrados de Médias Móveis - ARIMA( $p, d, q$ ) aos dados ([MORETTIN; TOLOI, 2018](#)).

### 2.3.1 Modelos Auto-regressivos de Ordem $p$

Um modelo auto-regressivo de ordem  $p$  - AR( $p$ ), é uma combinação linear de valores prévios do processo adicionado a um choque aleatório  $E_t$  ([BOX; JENKINS; REINSEL, 2008](#)).

Seja  $\tilde{U}_t = U_t - \mu$ , a série temporal subtraída do seu nível médio, então, um modelo AR( $p$ ) é dado por:

$$\tilde{U}_t = \phi_1 \tilde{U}_{t-1} + \phi_2 \tilde{U}_{t-2} + \dots + \phi_p \tilde{U}_{t-p} + E_t, \quad (2.5)$$

no qual os termos  $\phi_i$ ,  $i = 1, \dots, p$ , são os coeficientes Auto-regressivos.

Sendo o operador auto-regressivo de ordem  $p$  definido por:

$$\Phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - \dots - \phi_p(B^p) \quad (2.6)$$

e  $B$  o operador translação para o passado, definido como:

$$B^m U_t = U_{t-m}, \quad (2.7)$$

podemos escrever o modelo AR( $p$ ) como:

$$\Phi(B)\tilde{U}_t = E_t. \quad (2.8)$$

O processo será estacionário se  $\Phi(B) = \sum_{j=0}^{\infty} \phi^j B^j$  convergir para  $|B| \leq 1$  ([MORETTIN; TOLOI, 2018](#)).

### 2.3.2 Modelos de Médias Móveis de Ordem $q$

Um modelo de médias móveis - MA( $q$ ), é uma combinação linear de choques aleatórios  $E_t$ , dado por:

$$\tilde{U}_t = E_t - \theta_1 E_{t-1} - \theta_2 E_{t-2} - \dots - \theta_q E_{t-q}, \quad (2.9)$$

no qual os termos  $\theta_i$ ,  $i = 1, \dots, p$ , são os coeficientes de médias móveis.

Sendo o operador de médias móveis de ordem  $q$  definido por

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (2.10)$$

e  $B = B^m U_t = Z_{t-m}$ , o operador translação para o passado, podemos escrever o modelo MA( $q$ ) como:

$$\tilde{U}_t = \Theta(B)a_t. \quad (2.11)$$

### 2.3.3 Modelos Auto-regressivos de Médias Móveis

Para flexibilizar o ajuste do modelo à série pode-se utilizar o modelo auto-regressivo de médias móveis - ARMA( $p, q$ ), da forma

$$\tilde{U}_t = \phi_1 \tilde{U}_{t-1} + \phi_2 \tilde{U}_{t-2} + \dots + \phi_p \tilde{U}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}. \quad (2.12)$$

Sendo  $\Phi(B)$  e  $\Theta(B)$  os operadores auto-regressivo e de médias móveis, respectivamente, como definidos anteriormente, podemos escrever o processo ARMA( $p, q$ ) na forma compacta

$$\Phi(B) \tilde{U}_t = \Theta(B)E_t. \quad (2.13)$$

### 2.3.4 Modelos Auto-regressivos Integrados de Médias Móveis

A maioria das séries temporais reais apresentam alguma não estacionaridade, porém possuem comportamento homogêneo, não explosivo; para tais séries são tomadas diferenças que as tornam estacionárias. Para modelar adequadamente esse tipo de processo, [Box, Jenkins e Reinsel \(2008\)](#) indicam o modelo Auto Regressivo Integrado de Médias Móveis - ARIMA( $p, d, q$ ), definido por

$$\Phi(B)\Delta^d U_t = \Theta(B)E_t, \quad (2.14)$$

no qual  $p$  e  $q$  são as ordens de  $\Phi(B)$  e  $\Theta(B)$ , respectivamente e  $d$  é a ordem de diferenciação, na qual a  $n$ -ésima diferença é da forma

$$\Delta^n U_t = \Delta[\Delta^{n-1}U_t]. \quad (2.15)$$

A classe de modelos ARIMA é uma importante ferramenta de previsão e é a base de muitas ideias fundamentais na análise de séries temporais ([CHATFIELD, 2000](#)). Em todos os modelos vistos é necessário o ajuste dos parâmetros buscando o princípio da parcimônia, uso da menor quantidade de parâmetros ([BOX; JENKINS; REINSEL, 2008](#)).

É comum que séries temporais apresentem padrões repetitivos que aparecem regularmente em determinado intervalo de tempo, a sazonalidade. Para esse caso propõe-se

o modelo ARIMA Sazonal, o  $SARIMA(p, d, q)(P, D, Q)$ , representado da seguinte forma (BOX; JENKINS; REINSEL, 2008):

$$\Phi(B)\Phi(B^S)\Delta^d\Delta^D U_t = \Theta(B)\Theta(B^S)E_t, \quad (2.16)$$

no qual os valores  $P$ ,  $D$  e  $Q$  são as ordens dos polinômios auto-regressivo, de média móvel e do operador diferença sazonal, respectivamente;  $\Phi(B^S)$  é o polinômio auto-regressivo sazonal e  $\Theta(B^S)$  é o polinômio de médias móveis sazonal e  $\Delta^D$  o operador de diferença sazonal.

## 2.4 Modelos de Suavização Exponencial

Os modelos de suavização exponencial são bastante populares devido à sua simplicidade, eficiência computacional e razoável precisão (MORETTIN; TOLOI, 2018). Essa técnica consiste em atribuir pesos a observações passadas da série que decaiam exponencialmente, da mais recente a mais distante, ao longo do tempo, fazendo com que observações mais recentes tenham pesos maiores.

Este formalismo propõe a decomposição da série em suas componentes erro ( $E$ ), tendência ( $T$ ) e sazonalidade ( $S$ ), podendo ser classificados de acordo com o comportamento dessas componentes.

Dessa forma, uma série temporal  $u_t$  pode ser representada pela combinação dessas componentes nas formas aditiva (Equação 2.17) e multiplicativa (Equações 2.18 e 2.19).

$$u_t = T_t + S_t + E_t \quad (2.17)$$

$$u_t = T_t \times S_t \times E_t \quad (2.18)$$

$$u_t = (T_t + S_t) \times E_t \quad (2.19)$$

Segundo Barros et al. (2018), geralmente um modelo multiplicativo é apropriado quando a variação do padrão sazonal ou em torno da componente tendência for proporcional à magnitude dos dados, já o modelo aditivo é mais viável quando esta componente não for proporcional à magnitude dos dados.

A tendência pode se apresentar como nenhuma ( $N$ ), aditiva ( $A$ ), multiplicativa ( $M$ ), aditiva amortecida ( $Ad$ ) ou multiplicativa amortecida ( $Md$ ), já a sazonalidade pode ser nenhuma ( $N$ ), aditiva ( $A$ ) ou multiplicativa ( $M$ ). Na Tabela 1 são apresentadas 15

variações desse modelo, considerando ainda o erro como aditivo ou multiplicativo, tem-se um total de 30 modelos distintos (HYNDMAN; KOEHLER, 2008).

Um modelo ETS será representado aqui sob a notação  $ETS(E, T, S)$ , no qual  $E$  representa o Erro,  $T$  a tendência e  $S$  a sazonalidade.

Tabela 1 – Classificação ETS.

Tendência	Sazonalidade		
	N (Nenhuma)	A (Aditiva)	M (Multiplicativa)
N (Nenhuma)	NN	NA	NM
A (Aditiva)	AN	AA	AM
Ad (Aditiva Amortecida)	AdN	AdA	AdM
M (Multiplicativa)	MN	MA	MM
Md (Multiplicativa Amortecida)	MdN	MdA	MdM

Fonte: Hyndman et al. (2002).

Esse esquema de classificação foi proposto por McCormick (1969) e modificado por Gardner (1985), Hyndman et al. (2002), Taylor (2003) e Hyndman e Koehler (2008), que introduziu a notação ETS.

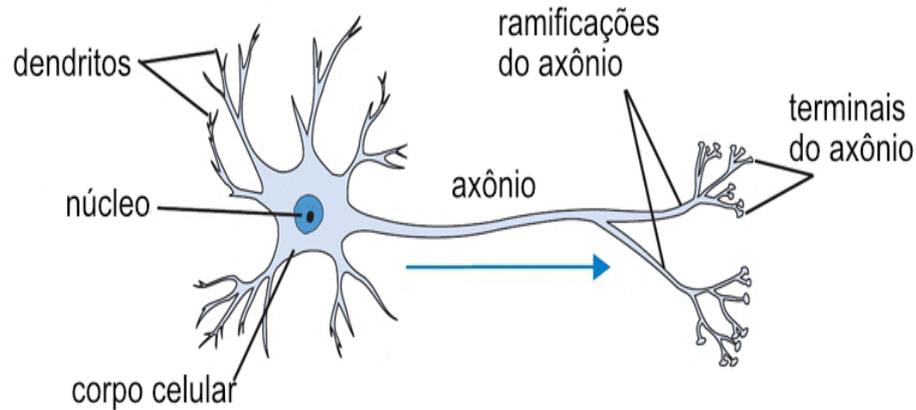
## 2.5 Redes Neurais Artificiais

O cérebro humano é uma das estruturas mais complexas e pouco compreendida. No entanto, o conhecimento básico da sua operação indica que é composto por, aproximadamente, dez mil milhões de unidades básicas, chamadas neurônios e cada uma dessas unidades conecta-se com outras dez mil. Um neurônio pode receber diversas entradas, que são adicionadas de alguma forma, caso receba entradas ativas suficientes de uma só vez, será ativado e emitirá um disparo, caso contrário, permanecerá inerte (BEALE; JACKSON, 1991). Uma ilustração do neurônio biológico e suas características básicas estão representados na Figura 1.

Inspirada no funcionamento do cérebro humano, uma Rede Neural Artificial (ANN) é um processador paralelo distribuído, constituído de unidades de processamento simples (neurônios ou nodos), que são naturalmente propensos a armazenarem conhecimento experimental e torná-lo disponível para o uso (HAYKIN, 2001).

Estas unidades são dispostas em um arranjo espacial composto de uma ou mais camadas e interligadas por um grande número de conexões. Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam a informação representada no

Figura 1 – Representação básica do Neurônio Biológico



Fonte: Bezerra (2016).

modelo e ponderam a entrada recebida por cada neurônio da rede (BRAGA; CARVALHO; LUDEMIR, 2007).

As redes neurais têm a capacidade de aprender por exemplos e generalizar o que aprenderam. A sua utilização na resolução de tarefas passa por uma fase de aprendizado, no caso um aprendizado supervisionado, quando a rede extrai informações relevantes dos padrões de informação a ela apresentados, o conjunto de treinamento, com entradas e saídas, e cria uma representação própria para o problema. Essa etapa consiste em um processo iterativo de ajuste dos parâmetros da rede, os pesos sinápticos, os quais guardam, ao final do processo, o conhecimento adquirido do ambiente no qual a rede está inserida, fazendo com que a rede consiga lidar com esse ambiente de maneira autônoma (HAYKIN, 2001) (BRAGA; CARVALHO; LUDEMIR, 2007).

Na literatura é possível encontrar diversos tipos de redes neurais, dentre estas, pode-se citar as Redes Neurais Convolucionárias (LECUN; BENGIO; HINTON, 2015); Redes Multilayer Perceptron (HAYKIN, 2001); Redes Recorrentes (HAYKIN, 2009) e Redes de Hopfield (HOPFIELD, 1982). As Redes Multilayer Perceptron, entretanto, destacam-se por sua

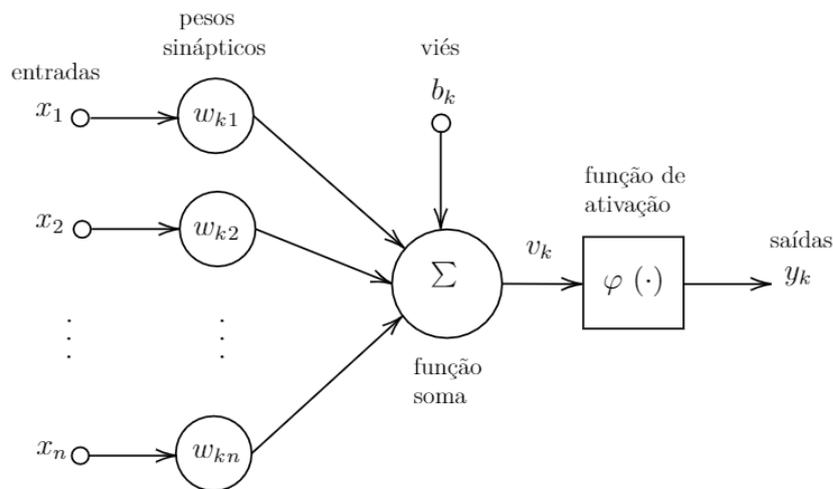
### 2.5.1 Rede Multilayer Perceptron – MLP

Em 1958, Frank Rosenblatt, propôs o modelo Perceptron, o modelo mais simples de ANN usado na classificação de padrões linearmente separáveis, baseado no neurônio de McCulloch–Pitts. A estrutura básica do Perceptron consiste em um único neurônio

com pesos sinápticos ajustáveis e o viés. Nesse caso, se os padrões utilizados para o treinamento do perceptron são extraídos de duas classes linearmente separáveis, o algoritmo do perceptron converge para um hiperplano separador entre as duas classes. Posteriormente, a demonstração desse fato ficou conhecida como o Teorema da Convergência Perceptron (HAYKIN, 2009).

Segundo Braga, Carvalho e Ludemir (2007), as redes MLP possuem poder computacional superior quando comparadas às redes sem camadas intermediárias, uma vez que estas conseguem lidar com problemas não linearmente separáveis. É consagrado na literatura que as estruturas MLP com uma única camada intermediária são capazes de aproximar qualquer função contínua e que duas camadas intermediárias são suficientes para aproximar qualquer função matemática. Para o caso de séries temporais, a MLP com uma camada oculta parece ser suficiente. Pode-se observar a representação de um neurônio artificial na Figura 2.

Figura 2 – Modelagem básica do Neurônio Artificial



Fonte: Braga; Carvalho; Ludemir (2007).

A unidade básica de processamento de informação do sistema nervoso central humano é o neurônio ( $\mathcal{N}$ ), composto por três elementos: um conjunto de dendritos ( $\mathcal{I}$ ), um corpo celular ( $\mathcal{S}$ ) e um axônio ( $\mathcal{A}$ ). Basicamente, o processo de uma ANN consiste em três ações realizadas pelos seus elementos básicos: 1) o conjunto de dendritos ( $\mathcal{I}$ ) recebe os estímulos externos ao corpo celular ( $\mathcal{S}$ ); 2)  $\mathcal{S}$  pondera esses estímulos através de operações agregativas simples e 3) o axônio ( $\mathcal{A}$ ) processa a informação recebida de  $\mathcal{S}$ , por meio de operações mais sofisticadas, gerando as respostas dos neurônios aos estímulos da camada de entrada. Naturalmente, as saídas de um neurônio podem ser as entradas de

outro, configurando assim uma rede neural. Geralmente, o neurônio  $\mathcal{N}$  é modelado por um Perceptron ( $\mathcal{P}$ ).

A abordagem para séries temporais com redes neurais, ou seja, a previsão para o valor da série  $u_t$ , contempla a aproximação

$$\mathcal{I}_t = \{u_{t-t_i}\}_{i \in \mathcal{I}_u}, \quad (2.20)$$

isto é, valores passados da série, e o modelo

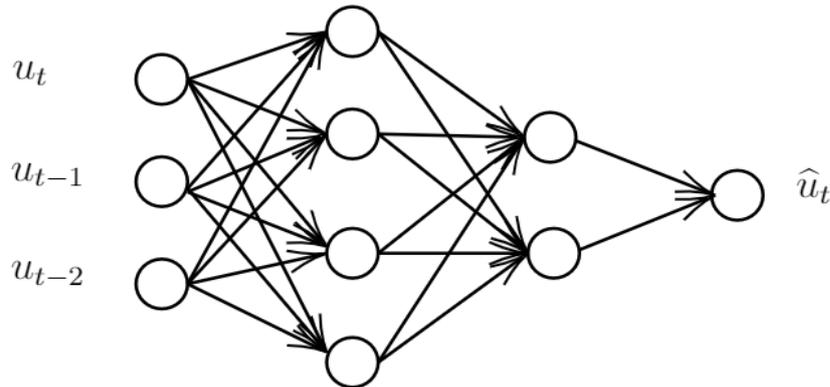
$$\mathcal{S}(\mathcal{I}_t) = \sum \theta_i u_{t-t_i} + \theta_0, \quad (2.21)$$

na qual  $\theta_i$  é o coeficiente que pondera as observações  $u_{t-t_i}$  e  $\theta_0$  é o intercepto. Dessa forma,  $\mathcal{S}$  envolve uma combinação linear dos dados de entrada. Em resposta a série regressa, obtém-se a previsão

$$\hat{u}_t = \mathcal{A}(\mathcal{S}(\mathcal{I}_t)), \quad (2.22)$$

em que a aproximação  $\mathcal{A}(\cdot)$  recebe o nome de Função de Ativação. A Figura 3 ilustra uma rede neural, com duas camadas intermediárias, para duas entradas regressas de uma série temporal.

Figura 3 – Arquitetura de Rede Neural adaptada para séries temporais com duas entradas regressas.



Fonte: Autor (2020).

Para séries temporais têm sido comum o uso de uma rede com uma camada intermediária, uma rede MLP, levando a uma função de soma simples

$$S_h(I_t) = \sum \theta_{hi} u_{t-t_i} + \theta_{h0}, \quad (2.23)$$

com funções de ativação intermediárias dadas por

$$A_h(S_h(I_t)), \quad (2.24)$$

em que  $h$  representa o índice do neurônio.

Por fim, as funções de ativação intermediárias são operadas por uma função de ativação final que se dedica à previsão dos valores futuros da série

$$\hat{u}_t = A_0 \left( \sum_{h=1}^H \theta_{h0} A_h(S_h(I_t)) \right). \quad (2.25)$$

### 2.5.2 Funções de Ativação

As funções de ativação têm o papel de informar às camadas anteriores os erros cometidos pela rede com a maior precisão possível (HAYKIN, 2001).

O cálculo para cada neurônio de uma rede MLP leva em consideração a derivada da função de ativação associada a esse neurônio. Para garantir a existência dessa derivada é necessário que a função de ativação seja contínua. Basicamente, diferenciabilidade é o único critério que uma função de ativação deve atender (HAYKIN, 2009). As funções de ativação mais comuns são a Linear, Sigmóide Logística e a Tangente Hiperbólica, cujos gráficos podem ser observados na Figura 4 e são descritas a seguir:

#### a) Função de Ativação Linear

$$y = \alpha x + \beta, \quad (2.26)$$

sendo  $\alpha \in \mathbb{R}$  define a saída linear para os valores de entrada  $x$ ,  $\beta$  é o intercepto e  $y$  é a saída;

#### b) Função de Ativação Sigmóide Logística

$$y = \frac{1}{1 + \exp(-\lambda x)}, \quad (2.27)$$

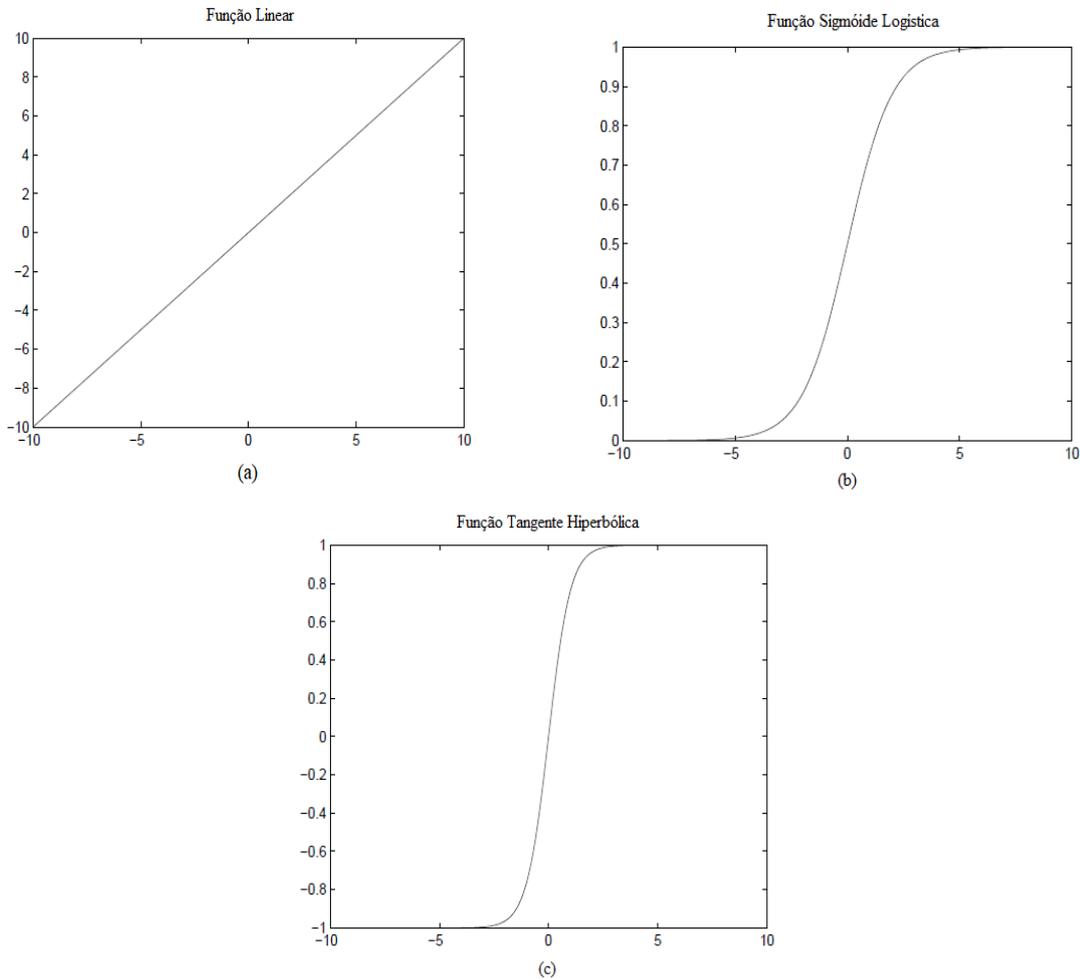
em que  $\lambda$  determina a sensibilidade de resposta da função. Geralmente ajusta-se  $\lambda = 1$ . Sua maior vantagem sobre a função linear é justamente a sua não linearidade, os valores dessa função variam entre 0 e 1.

#### c) Função de Ativação Tangente Hiperbólica]

$$y = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad (2.28)$$

esta função preserva a forma sigmóide da logística, porém passa a assumir valores no intervalo aberto  $(-1, 1)$ .

Figura 4 – Funções de Ativação Linear (a); Sigmóide Logística (b) e Tangente Hiperbólica (c)



Fonte: Autor (2020).

A notação aqui adotada para modelos de redes neurais será  $ANN(NI, NHL, NO)$ , com  $NI$  indicando o número de entradas auto-regressivas (auto-regressivas e auto-regressivas sazonais  $P_{AR} + P_{ARS}$ ),  $NHL$  é a quantidade de nodos na camada oculta e  $NO$  é o número de saídas, que, para a abordagem desse trabalho é sempre igual a 1, uma vez que trabalha-se com a previsão de um passo a frente.

### 2.5.3 Algoritmos de Aprendizado

Segundo Riedmiller (1994), um bom algoritmo de aprendizado deve atender pelo menos aos seguintes requisitos:

- i) Rápida convergência;

- ii) Fácil escolha dos parâmetros;
- iii) Boa capacidade de generalização para entradas desconhecidas.

O pacote para ANN em linguagem R ([COMPUTING, 2015](#)), *Neuralnet*, utiliza três tipos distintos de algoritmo: O *Bacpropagation*, o *Rpropagation* e aquele baseado no Algoritmo Globalmente Convergente.

O algoritmo *Backpropagation* ([MCCLELLAND; RUMELHART, 1986](#)) ou, em português, retroprogramação, é uma técnica baseada na descida do gradiente com a função Erro recebendo todos os pesos como argumentos e atribuindo-os ao erro de saída. Consiste na busca da minimização da função Erro através da descida do gradiente.

Esse algoritmo contém duas etapas: i) *Forward*, ou passo para frente e ii) *Backward*, passo para trás ([BRAGA; CARVALHO; LUDEMIR, 2007](#)). Na primeira etapa, a rede propaga o padrão de treinamento a partir das entradas para gerar uma saída. A saída correspondente é retroprogramada para gerar um erro, chamado delta ( $\Delta$ ), que consiste na diferença entre o valor real e o valor gerado pela rede. Na segunda etapa, os pesos são atualizados baseados no valor do respectivo delta, obtendo-se um valor mais próximo do real. Essas fases são repetidas até que o erro seja nulo ou até que seja inserido um ponto de parada na rede. A seguir é apresentado o pseudo-código do algoritmo *Backpropagation*.

---

**Algorithm 1** Pseudocódigo Algoritmo *Backpropagation*

---

```

begin
  Inicializar pesos e parâmetros;
  Para cada padrão de treinamento  $x$ :
    até o erro ser mínimo ou até realizar um dado número de ciclos
      Definir a saída da rede através da fase forward;
      Comparar as saídas produzidas com as desejadas;
      Atualizar os pesos dos nodos através da fase backward.
end

```

---

Existem limitações nessa técnica, além de demandar muitas interações até a convergência em superfícies mais complexas, a função pode atingir mínimos locais para o erro, causando um *underfitting* ([BRAGA; CARVALHO; LUDEMIR, 2007](#)), ([VAPNIK, 2000](#)).

O Algoritmo *RPropagation*, ou Retropragamação Resiliente, foi proposto por [Riedmiller e Braun \(1993\)](#), para superar as limitações da técnica baseada no gradiente descendente puro, *Backpropagation*. Nesse novo algoritmo, uma adaptação local das atualizações dos pesos é realizada, agora de acordo com o comportamento da função erro,

não levando em consideração a magnitude da derivada, mas apenas o seu sinal. Dessa forma, seu processo não é influenciado negativamente pelo tamanho imprevisível da derivada.

De acordo com o sinal da derivada a direção do ajuste dos pesos é definido (aumentar ou diminuir o peso anterior), o tamanho do ajuste é dado por um valor de atualização de acordo com a regra: se a derivada parcial do erro em relação a um peso mantém o sinal, o valor de atualização é aumentado por um fator  $\eta^+$ , acelerando a convergência. Caso a derivada mude o sinal, indicando que a última atualização foi muito grande e o algoritmo cedeu a um mínimo local, o valor da atualização é reduzido por um fator  $\eta^-$ . (BRAGA; CARVALHO; LUDEMIR, 2007).

O número de etapas de aprendizagem que este algoritmo requer é significativamente reduzido em detrimento de outros algoritmos e possui apenas cinco parâmetros, entretanto há um alto risco de overfitting. A primeira versão do algoritmo Rprop, implementa um conceito geral para melhorar o treinamento em rede denominado retorno de peso, o qual Igel e Hüsken (2003) nomearam de RPROP+.

Nesta variante, após a adaptação de cada peso, a atualização de pesos sugere que, se a derivada da função erro for positiva, o peso seja diminuído pelo seu valor atual e se for negativa, o valor atual seja a ele adicionado. Entretanto, quando a derivada mudar o seu sinal, a atualização do peso anterior é revertida, esse processo de retorno de pesos indica que a derivada deve alterar seu sinal mais uma vez na etapa seguinte, não havendo adaptação do valor de atualização na etapa posterior (RIEDMILLER; BRAUN, 1993).

Uma nova versão algoritmo Rprop foi apresentada por Riedmiller (1994), na qual o retorno de pesos é omitido e não há necessidade de armazenar o peso anterior das atualizações, Igel e Hüsken (2003) denotaram essa versão por RPROP-.

O Algoritmo Globalmente Convergente (GRPROP) é baseado na retroprogramação resiliente sem retorno de peso, neste é adicionada uma taxa de aprendizado. Apresenta duas variações: “SAG” na qual essa taxa de aprendizado associa o menor gradiente absoluto ou “SLR” na qual associa-se a menor taxa de aprendizado em si (COMPUTING, 2015).

## 2.6 Regressão de Vetor de Suporte

A técnica de Aprendizado de Máquina (AM) *Support Vector Machines* (SVM), do português Máquina de Vetores de Suporte vem sendo desenvolvida desde 1963, quando Lerner e Vapnik (1963) introduziram o algoritmo *Generalized Portrait*. As SVM's semelhantes a forma atualmente conhecida foram apresentadas por Boser, Guyon e Vapnik (1992), com uma generalização não linear do algoritmo *Generalized Portrait*. Essa técnica é utilizada em tarefas de classificação, reconhecimento de padrões e, para o caso de séries

temporais, análise de regressão.

Técnicas de AM se utilizam do princípio da indução para realizar inferências, através do qual é possível obter conclusões gerais a partir de um conjunto específico de exemplos. O aprendizado indutivo pode ocorrer de forma supervisionada, com a presença de um conjunto de exemplos com entrada e saída que permite o conhecimento do ambiente de trabalho, ou não-supervisionado, em que não existem dados rotulados (LORENA; CARVALHO, 2007).

O desenvolvimento de SVM trouxe uma nova perspectiva para a análise de dados não lineares, por possuir adaptabilidade superior a outros métodos (CHEN et al., 2004).

A forma alternativa de SVM, a *Support Vector Regression* (SVR), do português Regressão de Vetores de Suporte, apresentada inicialmente por Vapnik (1995), é empregada no contexto de previsão de séries temporais e tem por objetivo aproximar uma função usando os dados observados que “treina” a SVM (SMOLA; SCHOLKOPF, 2004; SAPANKEVYCH; SANKAR, 2009). A seguir é apresentada a teoria relativa a SVR.

Dado um conjunto de dados (amostra) de treinamento  $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^n \times \mathbb{R}$ , e que  $x_i$  é um vetor de entradas  $n$ -dimensional,  $y_i$  são as saídas e  $n$  é o número de observações no conjunto de treinamento (KANG; LI, 2016). A relação não linear entre entrada e saída pode ser definida como:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad (2.29)$$

onde  $f(\mathbf{x})$  é o valor predito com, no máximo, um desvio  $\epsilon$  de  $y_i$ ;  $b$  é o viés e  $\mathbf{w}$  é um vetor de pesos. Isso nos leva a um problema de otimização sujeito a restrições, em que o objetivo é encontrar um conjunto de pesos “ótimo” (SAPANKEVYCH; SANKAR, 2009), dado por

$$\text{minimizar} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n [L(\xi_i) + L(\xi_i^*)], \quad (2.30)$$

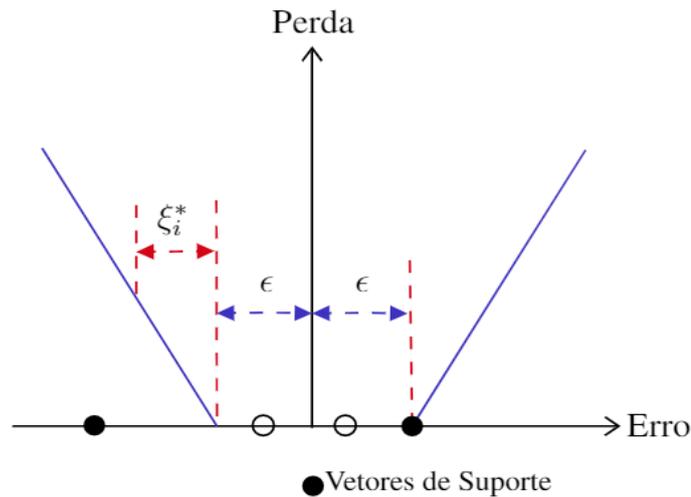
$$\text{sujeito a} \quad \begin{cases} y_i - f(x_i) \leq \epsilon + \xi_i \\ f(x_i) - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \end{cases} \quad (2.31)$$

na qual  $C > 0$  é a constante de regularização que indica a quantidade de erros no conjunto de treinamento,  $\|\cdot\|$  é a norma do vetor e  $L(\cdot)$  é a função de perda  $\epsilon$ -insensível (Figura 5),

dada por:

$$L(\xi) = \begin{cases} 0 & \text{se } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{caso contrário .} \end{cases} \quad (2.32)$$

Figura 5 – Função Perda  $\epsilon$ -insensível.



Fonte: Kang; Li (2016).

Aplicando o método dos Multiplicadores de Lagrange ao problema de otimização anterior, temos a seguinte expansão:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{w}^T \mathbf{x} + b, \quad (2.33)$$

em que  $\alpha_i$  e  $\alpha_i^*$  são os multiplicadores de Lagrange.

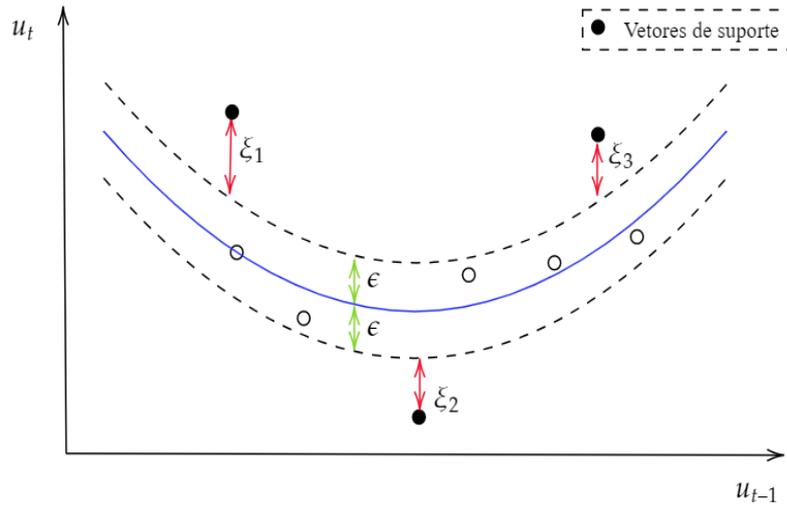
Quando os dados não são lineares em seu espaço de entrada é necessário estender essa regressão para um caso de regressão não linear, para tal é inserida uma função Kernel, que mapeia os dados de entrada para um espaço de dimensão superior onde a regressão linear torna-se possível. Portanto, teremos:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}_j) + b. \quad (2.34)$$

Apenas para valores  $|f(x_i) - y_i| \geq \epsilon$  os multiplicadores de Lagrange,  $\alpha_i$  e  $\alpha_i^*$ , podem ser diferentes de zero, isto é, todas as amostras que estão no  $\epsilon$ -tubo (região pontilhada da Figura 6) possuem multiplicadores de Lagrange nulos. Os pontos com coeficientes diferentes de zero são chamados de vetores de suporte (SMOLA; SCHOLKOPF, 2004).

A Figura 6 traz uma ilustração para SVR aplicado ao caso de séries temporais.

Figura 6 – SVR para séries temporais.



Fonte: Autor (2020).

### 2.6.1 Funções Kernel

Considerando  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^{n+h}$ , com  $\mathbf{x} \rightarrow \phi(\mathbf{x})$ , define-se o Kernel conforme a equação abaixo:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j). \quad (2.35)$$

Segundo Vapnik (2000), uma função para ser Kernel precisa apresentar duas características:

*i)* Ser simétrica:

$$K(x, z) = (\phi(x) \cdot \phi(z)) = K(z, x);$$

*ii)* Satisfazer as equações de Cauchy-Schwarz:

$$K(x, z)^2 = (\phi(x) \cdot \phi(z))^2 \leq \|\phi(x)\|^2 \cdot \|\phi(z)\|^2 = (\phi(x) \cdot \phi(x))(\phi(z) \cdot \phi(z)) = K(x, x) \cdot K(z, z).$$

Entretanto, essas condições não são suficientes para garantir a existência do espaço de características. Dessa forma, uma Função Kernel deve ainda satisfazer o Teorema de Mercer.

**Teorema 2.1** (Teorema de Mercer). *Para garantir que a função simétrica*

$$K(x, z) = \sum_{i=1}^{\infty} \phi(x)_i \cdot \phi(z)_i \quad (2.36)$$

*seja um Kernel, é necessário que*

$$\int \int K(x, z) g(x) g(z) dx dz > 0 \quad (2.37)$$

*seja válida para todo  $g \neq 0$  e*

$$\int g^2(x) dx < \infty. \quad (2.38)$$

As funções Kernel retornam o produto interno entre dois pontos no espaço de entrada. A expressão para a função Kernel pode ser escolhida de forma direta sem que a função  $\Phi(\cdot)$  esteja explícita.

A quantidade de parâmetros existentes na função Kernel afeta muito a complexidade dos modelos (SU et al., 2013). As Funções Kernel mais comuns na literatura são Linear, Polinomial, Base Radial e Sigmóide.

#### Kernel Linear

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \quad (2.39)$$

Essa Kernel não possui nenhum parâmetro próprio para ajustar.

#### Kernel Polinomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + coef0)^{degree}, \quad (2.40)$$

o parâmetro  $degree > 0$  representa o grau do polinômio e o  $coef0$  é o intercepto, ambos necessitam de ajuste.

#### Kernel de Base Radial

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \langle \mathbf{x}_i - \mathbf{x}_j, \mathbf{x}_i - \mathbf{x}_j \rangle^2). \quad (2.41)$$

neste Kernel,  $\gamma = 1/2\sigma^2$  com  $\sigma > 0$  que precisa ser ajustado.

## Kernel Sigmóide

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \langle \mathbf{x}_i, \mathbf{x}_j \rangle + coef0), \quad (2.42)$$

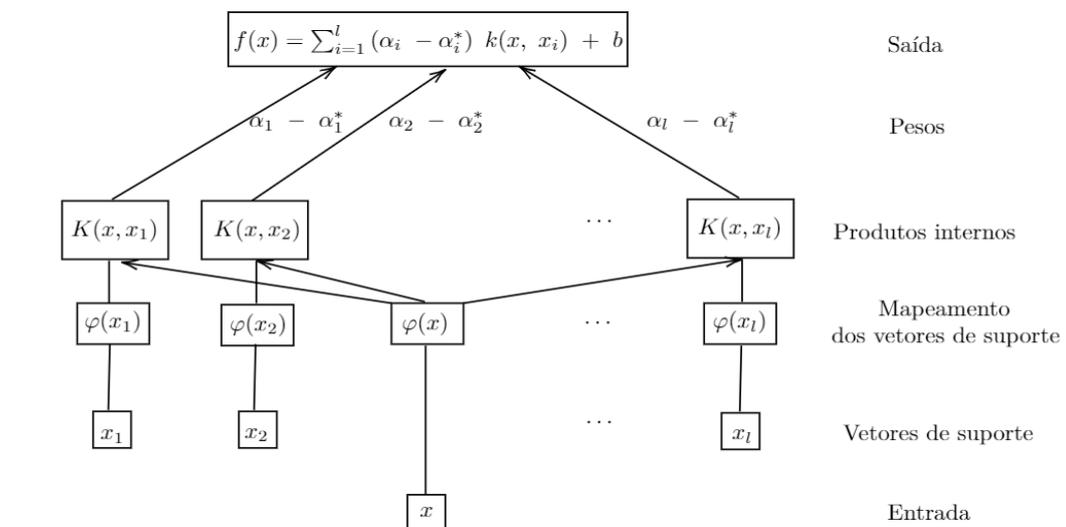
A utilização de Funções Kernel reduz o número de parâmetros livres na computação de problemas envolvendo dados não lineares, fato que confere à SVM superioridade enquanto técnica de AM (CHEN et al., 2004).

Os parâmetros da SVR exercem bastante influência nos resultados da previsão, uma escolha equivocada de parâmetros pode levar a um *overfitting* (super ajuste) ou *underfitting* (sub ajuste). Dessa forma, a seleção dos hiperparâmetros é uma tarefa muito importante da modelagem com SVR (KANG; LI, 2016).

O desempenho da SVR está ligado à combinação dos parâmetros  $C$ ,  $\epsilon$  e o(s) parâmetro(s) da função Kernel. O parâmetro  $C$ , chamado de penalização, é responsável pelo *tradeoff* entre a maximização da margem do  $\epsilon$ -tubo, quando são aceitos valores com desvios maiores que  $\epsilon$ , e a minimização do erro no conjunto de treinamento. Se o seu valor for muito alto ou muito baixo pode haver *overfitting* ou *underfitting* (WANG et al., 2014). O parâmetro  $\epsilon$  é responsável pelo raio do  $\epsilon$ -tubo ao redor da função de regressão, considera-se também o desvio permitidos a previsão em relação aos dados alvos.

Após a identificação dos parâmetros, o modelo pode ser utilizado para prever a saída correspondente a uma entrada desconhecida. Uma arquitetura da SVR é apresentada na Figura 7.

Figura 7 – Arquitetura do processo de treinamento para SVR.



Fonte: Kang; Li (2016).

Neste esquema, a entrada  $x$  e os vetores de suporte  $x_i$  são mapeados não linearmente (por  $\varphi$ ) em um espaço de características, no qual os produtos internos são calculados através da Função Kernel. Na prática, esses dois passos são computados de uma única vez. Os resultados são linearmente combinados pelos pesos  $\alpha_i$ , encontrados através da solução de um problema quadrático e, finalmente, essa combinação linear é reunida pela função  $f(\cdot)$ .

Uma tarefa muito difícil na abordagem de SVR é a estimação ótima dos seus parâmetros e dos parâmetros da função Kernel, uma vez que não há um consenso entre os pesquisadores da área quanto ao melhor método pra esse fim.

Na literatura é possível encontrar diversos algoritmos utilizados na otimização dos parâmetros de SVR, por exemplo, o Algoritmo Genético (GA), uma metaheurística inspirada no processo de seleção natural (SAPANKEVYCH; SANKAR, 2009) ou Otimização por Enxame de Partículas (PSO) como nos trabalhos de Nieto et al. (2015), Wang et al. (2014) e Dong et al. (2015).

Há ainda trabalhos como o de Ren e Bai (2010) que aplicaram e compararam os métodos GA e PSO apresentando resultados satisfatórios para o problema de estimação dos parâmetros de SVM a um razoável custo computacional, sobretudo para o PSO.

Entretanto, segundo Chen et al. (2004) as abordagens práticas que existem para a selecionar os hiperparâmetros de SVR possuem algumas deficiências. Neste trabalho, optou-se por uma abordagem distinta, agora baseada no Algoritmo Simulated Annealing Kirkpatrick, Gelatt e Vecchi (1983), também utilizado para otimização das Redes Neurais aqui trabalhadas.

## 2.7 Formalismos de Modelagem Combinada

Os estudos realizados até aqui apontam a combinação de preditores como alternativa de modelagem que vem produzindo resultados estatisticamente atraentes, autores como Bates e Granger (1969), Patton (2012) e Firmino, Neto e Ferreira (2014) contribuem com essa assertiva.

Na literatura é possível encontrar diversos trabalhos que apresentam a combinação de preditores como uma alternativa para previsão de séries temporais, Jaganathan e Prakash (2020) utilizaram a combinação via média e mediana para séries financeiras com diferentes frequências; os autores Petropoulos e Svetunkov (2020) combinaram quatro modelos univariados para séries simuladas através da mediana, obtendo bons resultados; já Oliveira et al. (2016) mostraram a utilidade do método de combinação via cópulas quando comparado à média simples e mínima variância para séries simuladas e reais.

Outros estudos utilizam ainda séries temporais *fuzzy* e redes neurais convolucionais para previsão de carga (SADAEI et al., 2019) e combinação ARIMA-SVR baseado na análise de Wavelets para previsão de séries hidrológicas (XIE; LOU, 2019).

Para a modelagem combinada, considera-se agora as estimativas geradas pelos  $k$  modelos individuais,  $\hat{u}_t$ , dos formalismos citados,  $\{\hat{u}_{ti}\}_{i=1}^k$ ,  $t = 1, \dots, k$ . Para os objetivos deste trabalho serão considerados quatro tipos de combinadores: Média Simples (cSA), Mediana Simples (cSM), Cópula (cCB) e Regressão de Vetor de Suporte (cSVR).

### 2.7.1 Média Simples e Mediana Simples

A predição combinada para  $u_t$  via Média Simples, aqui chamada cSA, é uma combinação linear que calcula basicamente a média aritmética simples dos  $k$  modelos individuais:

$$\hat{u}_t^{SA} = \frac{\sum_{i=1}^k \hat{u}_{t,i}}{k}, \quad (2.43)$$

nesse caso,  $\hat{u}_t^{SA}$  é a previsão por Média Simples dos  $k$  modelos individuais e  $\hat{u}_{t,i}$  é a previsão do  $i$ -ésimo modelo para o valor  $u_t$ .

A combinação por Mediana Simples, aqui chamada cSM, se dá através da mediana simples dos  $k$  preditores individuais. Chama-se de mediana o valor que divide um conjunto de dados ordenados em duas metades.

### 2.7.2 Cópulas

A princípio são fornecidos conceitos relevantes acerca de cópulas e em seguida apresenta-se os formalismos da combinação via cópula, aqui chamada cCB.

A teoria de cópulas foi introduzida inicialmente por Sklar (1959) e se destacou na modelagem multivariada por incorporar a estrutura de dependência entre as distribuições univariadas. Cópulas são capazes de combinar preditores de séries temporais independentemente das distribuições marginais e conjunta, tornando-se um método competitivo no campo. Formalmente, infere-se a distribuição marginal do erro de cada preditor e, em seguida, a distribuição conjunta.

O formalismo de cópulas é fundamentado nos estudos de Sklar, a luz do Teorema que leva seu nome, o qual aborda e esclarece a relação de cópulas com as funções de distribuições marginais e as multivariadas (NELSEN, 2013).

Seja um vetor aleatório de  $k$  entradas  $\mathbf{X} = (X_1, \dots, X_k)$ , o teorema supracitado mostra que existe uma função acumulada conjunta,  $H_{X_1, \dots, X_k}(x_1, \dots, x_k)$ , com distribuições

acumuladas marginais  $v_i = F_{X_i}(x_i)$ , de forma que

$$H_{X_1, \dots, X_k}(x_1, \dots, x_k) = C(v_1, \dots, v_k). \quad (2.44)$$

Então,  $C(v_i)$  é uma cópula e a variável aleatória  $V_i \sim U[0, 1]$ . Se  $F_{X_1}, \dots, F_{X_k}$  são contínuas, [Bhatti e Do \(2019\)](#) afirma que a cópula  $C(\cdot)$  associada a  $H(\cdot)$  é única e pode ser obtida por:

$$C(v_1, \dots, v_k) = H\left(F_{X_1}^{-1}(v_1), \dots, F_{X_k}^{-1}(v_k)\right). \quad (2.45)$$

Aliando-se isso ao fato de que toda variável aleatória contínua pode ser transformada, através da sua função de distribuição acumulada marginal, em uma uniforme no intervalo  $[0, 1]$ , cópulas podem ser usadas para fornecer uma estrutura de dependência multivariada separadamente das distribuições marginais ([YAN, 2007](#)). Formalmente, cópulas são funções de distribuição multivariada cujas distribuições marginais univariadas possuem distribuição Uniforme no intervalo  $[0, 1]$  ([NELSEN, 2013](#)).

A FDP conjunta de  $\mathbf{X}$ ,  $p_{X_1, \dots, X_k}(x_1, \dots, x_k)$  pode ser obtida pela diferenciação das  $F_1, \dots, F_k$ :

$$p_{X_1, \dots, X_k}(x_1, \dots, x_k) = \frac{\partial^k F_{X_1, \dots, X_k}(x_1, \dots, x_k)}{\partial x_1 \dots \partial x_k} = c(F_{X_1}(x_1), \dots, F_{X_k}(x_k)) \prod_{i=1}^k p_{X_i}(x_i), \quad (2.46)$$

em que  $p_{X_i}(x_i)$  é a função densidade de probabilidade associada a  $X_i$  e  $c(\cdot)$  é a função densidade de probabilidade da cópula que pode ser obtida pela diferenciação de [2.44](#):

$$c(v_1, \dots, v_k) = \frac{\partial^k}{\partial v_1 \dots \partial v_k} C(v_1, \dots, v_k). \quad (2.47)$$

## Famílias de Cópulas

Na literatura é possível encontrar diversos tipos de cópulas ([NELSEN, 2013](#)). Para os propósitos deste trabalho serão apresentadas cópulas Arquimedianas e Elípticas.

### Cópulas Arquimedianas

As Cópulas Arquimedianas (CA) são populares devido a sua facilidade de derivação e capacidade de captar amplas faixas de dependência ([TRIVEDI; ZIMMER, 2011](#)).

Além disso, esta família de cópulas possui forma simples, fácil construção e propriedades matemáticas desejáveis (NELSEN, 2013).

Uma CA é definida da seguinte forma:

$$C(v_1, \dots, v_k) = \varphi^{-1}(\varphi(v_1) + \dots + \varphi(v_k)), \quad (2.48)$$

em que  $\varphi(\cdot)$  é chamada de função geradora da cópula e  $\varphi^{-1}(\cdot)$  é a sua inversa.

Dentre as cópulas da família Arquimediana serão destacadas as cópulas Clayton, Frank, Gumbel e Joe.

### Cópula Clayton

A cópula Clayton é assimétrica e tem função geradora dada por:

$$\varphi(t) = \frac{1}{\theta(t^{-\theta} - 1)}. \quad (2.49)$$

Substituindo a Equação 2.49 na Equação 2.48, obtém-se a cópula Clayton na forma

$$C(v_1, \dots, v_k) = \max\left((v_1^{-\theta} + \dots + v_k^{-\theta} + n - 1)^{-\frac{1}{\theta}}, 0\right), \quad (2.50)$$

em que  $\theta$  é o parâmetro de dependência e está restrito à região  $(0, \infty)$ . Quando  $\theta$  tende a zero, as marginais tendem a se tornarem independentes. Essa cópula não é capaz de explicar dependência negativa (TRIVEDI; ZIMMER, 2011).

### Cópula Frank

A cópula Frank tem função geradora dada pela fórmula

$$\varphi(t) = -\ln\left(\frac{\exp(-\theta t) - 1}{\exp(-t) - 1}\right). \quad (2.51)$$

Substituindo a Equação 2.51 na Equação 2.48, obtém-se a cópula Frank na forma

$$C(v_1, \dots, v_k) = -\ln\left(1 + \frac{(\exp(-\theta v_1) - 1) \dots (\exp(-\theta v_k) - 1)}{\exp(-\theta) - 1}\right), \quad (2.52)$$

com  $\theta > 0$  sendo o parâmetro que mede o grau de dependência entre as variáveis. Essa cópula tem propriedades idênticas à cópula Normal (CHERUBINI et al., 2012).

### Cópula Gumbel

A função geradora da cópula Gumbel é dada por

$$\varphi(v_i) = (-\ln v_i)^\theta, \quad (2.53)$$

com  $\theta \geq 1$ , quando substitui-se a Equação 2.53 na Equação 2.48 obtém-se a cópula Gumbel

$$C(v_1, \dots, v_k) = \exp\left(-\left((-\ln v_1)^\theta + \dots + (-\ln v_k)^\theta\right)^{\frac{1}{\theta}}\right). \quad (2.54)$$

A cópula Gumbel se limita apenas a dependência positiva (TRIVEDI; ZIMMER, 2011).

### Cópula Joe

A função geradora da cópula Joe é dada por

$$\varphi(t) = -\ln(1 - (1 - t)^\theta), \quad (2.55)$$

e assume a forma

$$C(v_1, \dots, v_k) = 1 - \left[(1 - v_1)^\theta + \dots + (1 - v_k)^\theta - (1 - v_1)^\theta \dots (1 - v_k)^\theta\right]^{1/\theta}, \quad (2.56)$$

com  $1 \leq \theta < \infty$ . De forma similar à Gumbel, esta cópula não explica a dependência negativa. É radialmente assimétrica com fraca dependência da cauda esquerda e forte dependência da cauda direita (BHAT; ELURU, 2009).

### Cópiulas Elípticas

As Cópiulas Elípticas (CE) não apresentam uma representação explícita devido a sua formulação ser dada implicitamente pela equação 2.45. As CE não são capazes de capturar assimetrias radiais e comportamento assimétrico da cauda (KUROWICKA; JOE, 2011). São consideradas aqui as cópiulas Normal ou Gaussiana e a *t* de *Student*.

### Cópula Normal

A cópula Normal, também conhecida como Gaussiana, é simétrica e modela a dependência por uma matriz definida positiva, em que os elementos descrevem a dependência entre as variáveis duas a duas. Sua utilização é conveniente quando se trabalha com duas ou mais dimensões e sua principal vantagem é a simplicidade (RENARD; LANG, 2007).

Considere  $\Phi(\cdot)$  a distribuição normal padrão multivariada na  $n$ -ésima dimensão com correlação dada pela matriz  $n$ -dimensional  $\rho$  (CHERUBINI et al., 2012). Assim, a cópula Normal pode ser escrita na forma

$$C(v_1, \dots, v_k) = \Phi(\varphi^{-1}(v_1), \dots, \varphi^{-1}(v_k) | \rho), \quad (2.57)$$

em que  $\varphi^{-1}(\cdot)$  é a inversa da distribuição normal padrão univariada.

A função densidade da cópula Normal é dada por:

$$\begin{aligned} c(v_1, \dots, v_k) &= c(\varphi^{-1}(v_1), \dots, \varphi^{-1}(v_k) | \rho) \\ &= \frac{1}{\sqrt{|\rho|}} \exp \left( -\frac{1}{2} (\varphi^{-1}(v_1), \dots, \varphi^{-1}(v_k)) (\rho^{-1} - \mathbb{I}) \begin{pmatrix} \varphi^{-1}(v_1) \\ \vdots \\ \varphi^{-1}(v_k) \end{pmatrix} \right) \end{aligned} \quad (2.58)$$

com  $\mathbb{I}$  indicando a matriz identidade  $n$ -dimensional.

## Distribuições de Probabilidade Marginais

Neste trabalho foram utilizadas com variações de Distribuições Marginais para a combinação via cópula as seguintes distribuições: Normal, Normal-Assimétrica, Laplace e Cauchy, cujas principais características são apresentadas a seguir.

### Distribuição Normal

A variável aleatória  $X$  segue o modelo Normal se sua densidade de probabilidade é da forma (MAGALHÃES, 2015):

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] I_{(-\infty, \infty)}(x), \quad \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^*. \quad (2.59)$$

Os parâmetros  $\mu$  e  $\sigma^2$  representam, respectivamente, a média e a variância da variável. Usa-se a notação  $X \sim N(\mu, \sigma)$  para indicar que uma variável pode ser modelada por uma distribuição Normal. É uma distribuição simétrica e em formato de sino.

Sendo  $X$  como definida acima e seja  $Z = (X - \mu)/\sigma$  uma variável aleatória. Então  $Z$  tem uma distribuição  $N(0, 1)$ , conhecida como Normal-Padrão, que tem função densidade de probabilidade dada por:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{z^2}{2} \right). \quad (2.60)$$

Neste caso,  $\mu = 1$  e  $\sigma = 0$ .

### Distribuição Normal-Assimétrica

A variável aleatória  $X$  segue o modelo Normal-assimétrica, ou Skew-Normal, se sua densidade de probabilidade é da forma (AZZALINI, 1985):

$$f(x) = 2\varphi(x)\Phi(\lambda x), \quad (2.61)$$

em que  $\lambda \in \mathbb{R}$  define o parâmetro de assimetria e  $\varphi(\cdot)$  e  $\Phi(\cdot)$  denotam, nessa ordem, a função densidade de probabilidade e a distribuição acumulada da distribuição Normal-padrão. O valor  $\lambda > 0$  indica assimetria positiva e  $\lambda < 0$ , assimetria negativa.

### Distribuição de Laplace

A variável aleatória  $X$  segue o modelo de Laplace, ou exponencial dupla se sua densidade de probabilidade é da forma (KOTZ; KOZUBOWSKI; PODGORSKI, 2012):

$$f(x) = \frac{2}{\sigma} \exp\left(-\frac{|x - \theta|}{s}\right), \quad -\infty < x < \infty, \quad -\infty < \theta < \infty, \quad s > 0 \quad (2.62)$$

em que,  $\theta$  e  $s$  representam os parâmetros de localização e escala, respectivamente. A clássica distribuição de Laplace é simétrica.

### Distribuição de Cauchy

A variável aleatória  $X$  segue o modelo de Cauchy se sua densidade de probabilidade é da forma (CASELLA; BERGER, 2002):

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty, \quad (2.63)$$

O parâmetro  $\theta$  indica a mediana. Essa distribuição é simétrica e em forma de sino, porém diferentemente da Normal, não possui média.

## 2.7.3 Combinação via Cópulas

Para a combinação via cópula considere  $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,i}, \dots, X_{t,k})$  o vetor de  $k$  preditores individuais para as observações  $u_t$  da série temporal e  $\mathbf{x}_t = (x_{t,1}, \dots, x_{t,i}, \dots, x_{t,k})$  o vetor das estimativas correspondentes aos preditores de  $u_t$ . Seja  $\mathbf{E}_t$  o vetor de erros dos

preditores individuais, dado por  $\mathbf{E}_t = (E_{t,1}, \dots, E_{t,i}, \dots, E_{t,k})$ . Para o caso de erros aditivos, tem-se

$$E_{t,i} = u_t - X_{t,i}. \quad (2.64)$$

Assim, diante da estimativa para  $u_t$ ,  $x_{t,i}$ , o resíduo (a instância) associado a  $E_{t,i}$  pode se dar por

$$e_{t,i} = u_t - x_{t,i}. \quad (2.65)$$

Os casos de erros multiplicativos são semelhantes, contudo substituindo a subtração (“-”) pela divisão (“/”). Logo,  $\mathbf{e}_t = (e_{t,1}, \dots, e_{t,i}, \dots, e_{t,k})$  é o conjunto de resíduos observados das estimativas para  $u_t$ , isto é, uma instância de  $\mathbf{E}_t$ .

A distribuição acumulada conjunta de  $\mathbf{E}_{t,i}$  é dada por  $F_{\mathbf{E}_{t,i}}(e_{t,1}, \dots, e_{t,k})$  com distribuição acumulada marginal  $F_{E_{t,i}}(e_{t,i})$ . Para o caso de erros aditivos, tem-se que a distribuição acumulada marginal de  $E_{t,i}$  será

$$F_{E_{t,i}}(e_{t,i}) = P(E_{t,i} < u_t - x_{t,i}), \quad (2.66)$$

com FDP

$$p_{E_{t,i}}(e_{t,i}) = p_{E_{t,i}}(u_t - x_{t,i}). \quad (2.67)$$

Sabe-se que uma cópula é uma Função de Distribuição Acumulada (FDA) com marginais seguindo uma Uniforme em  $[0, 1]$ , sendo, então,  $v_i (i = 1, \dots, k)$  uma instância de  $k$  distribuições marginais univariadas com  $v_i \in [0, 1]$ . A cópula é dada pela FDA conjunta

$$C(v_1, \dots, v_k), \quad (2.68)$$

sendo  $v_i$  o valor da FDA dos erros do modelo  $X_i$  diante do resíduo  $e_{t,i}$ :

$$v_i = F_{E_{t,i}}(e_{t,i} | \alpha_i), \quad (2.69)$$

com  $\alpha_i$  representando o conjunto de parâmetros (ou vetor de parâmetros) da FDA marginal de  $E_i$ . Note-se então que as cópulas, através da Equação 2.69, transformam  $(e_{t,1}, \dots, e_{t,i}, \dots, e_{t,k})$  em  $(v_1, \dots, v_i, \dots, v_k)$ , assim, a Função Densidade de Probabilidade (FDP) conjunta de  $\mathbf{E}_t$  será

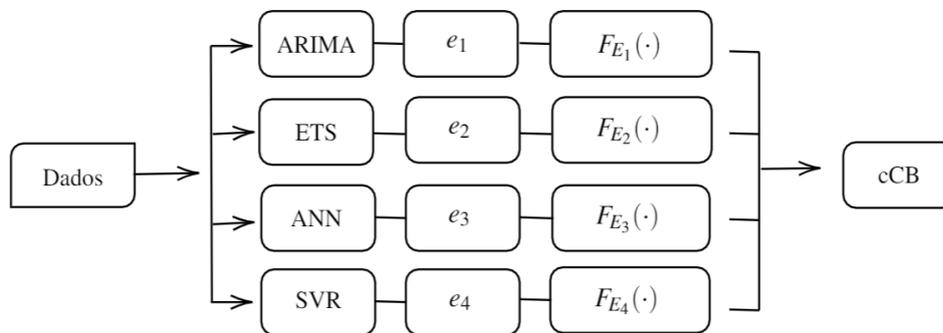
$$p_{\mathbf{E}_t}(e_{t,1}, \dots, e_{t,k}) = c_{\mathbf{E}_t}(v_1, \dots, v_k) \prod_{i=1}^k p_{E_{t,i}}(e_{t,i}). \quad (2.70)$$

Vale destacar que, diante do conjunto de estimativas para  $u_t$ , provenientes dos  $k$  modelos  $(x_{t,1}, \dots, x_{t,k})$ , a única incógnita da Equação (2.70) passa a ser  $u_t$ . Isto permite o uso de estimativas de máxima verossimilhança que levam em consideração a distribuição

dos erros dos modelos ao prever  $u_t$ , tal como realizado no presente trabalho. Mais especificamente, o espaço de possibilidades de  $u_t$  é representado por 1000 pontos igualmente espaçados e, deles, o valor mais verossímil para  $u_t$ , de acordo com a Equação (2.70), é adotado como estimativa do combinador baseado em cópulas (cCB).

Aqui, supõem-se que  $c_{\mathbf{E}_t}$  e  $p_{E_{t,i}}(e_{t,i})$  são invariantes no tempo, podendo-se suprimir o índice  $t$ . Há ainda que se comentar que  $C(\cdot)$  pode ser qualquer cópula. A Figura 8 mostra uma arquitetura simplificada da modelagem combinada via cópula.

Figura 8 – Arquitetura da Combinação via Cópula.



Fonte: Autor (2020).

### Inferência dos parâmetros das distribuições marginais e da cópula

Para inferir os parâmetros das distribuições marginais e da cópula adotou-se o Método da Máxima Verossimilhança (MMV), que, segundo Casella e Berger (2002), é, de longe, a técnica mais popular para determinar estimadores.

A função de Verossimilhança adaptada para os propósitos desse estudo é dada por

$$L = \prod_{t=1}^n p_{E_1, \dots, E_k}(e_{t,1}, \dots, e_{t,i}, \dots, e_{t,k} | \alpha_1, \dots, \alpha_i, \dots, \alpha_k, \gamma) \quad (2.71)$$

em que  $p_{\mathbf{E}_t}(\cdot)$  é a FDP conjunta dos erros  $\mathbf{E}_t$ ,  $\alpha_i$  é o vetor de parâmetros da distribuição marginal de  $E_i$  e  $\gamma$  é o vetor de parâmetros da cópula.

O Estimador de Máxima Verossimilhança (EMV) é o valor que maximiza a Equação 2.71, para maior conveniência de cálculos, trabalha-se com o logaritmo natural, dessa forma, maximizar 2.71 é equivalente a maximizar a chamada log-verossimilhança, expressa

na Equação 2.72:

$$l = \sum_{t=1}^n \log p_{E_1, \dots, E_k}(e_{t,1}, \dots, e_{t,i}, \dots, e_{t,k} | \alpha_1, \dots, \alpha_i, \dots, \alpha_k, \gamma). \quad (2.72)$$

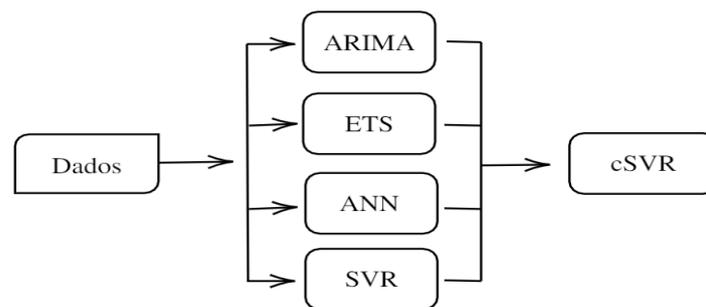
Essa alternativa é possível porque o logaritmo natural da função de verossimilhança é estritamente crescente no intervalo  $(0, \infty)$ , implicando que os extremos da Equações 2.71 e 2.72 coincidem (CASELLA; BERGER, 2002).

O MMV, entretanto, pode ser muito custoso computacionalmente, devido a estimação conjunta dos parâmetros da cópula e das distribuições marginais. Para solucionar esse problema, utiliza-se o Método IFM - do inglês *Inference for Marginals* (JOE; XU, 1996). Essa abordagem consiste em estimar separadamente os parâmetros das marginais e, em seguida, estimar os parâmetros da cópula condicionado as estimativas obtidas para as marginais, ambos pelo MMV.

#### 2.7.4 Combinação via *Support Vector Regression*

Para a combinação baseada em *Support Vector Regression*, aqui intitulada cSVR, serão tomadas como entradas da máquina de aprendizado as previsões dos modelos individuais (ARIMA, ETS, ANN e SVR) e os alvos serão os dados observados. A Figura 9 ilustra esse processo.

Figura 9 – Arquitetura da Combinação via SVR.



Fonte: Autor (2020).

De forma semelhante à modelagem SVR individual, os parâmetros serão otimizados através do Algoritmo *Simulated Annealing*.

## 2.8 Algoritmo *Simulated Annealing*

O Algoritmo *Simulated Annealing* (SA), Recozimento Simulado, em tradução livre, foi desenvolvido em 1983 por Kirkpatrick, Gelatt e Vecchi (1983) no âmbito da mecânica

estatística. Trata-se de uma poderosa técnica para encontrar soluções quase perfeitas para problemas de otimização combinatória. Baseado nos conceitos da Termodinâmica, o SA simula computacionalmente o processo de recozimento do metal, utilizado para fundir metais, no qual o metal é aquecido a uma temperatura elevada e resfriado lentamente de forma a obter maior homogeneidade no produto final (WHITE, 1984).

O pseudocódigo a seguir apresenta o processo simplificado do SA, no qual a busca é iniciada por uma solução inicial qualquer,  $p_0$ , a uma temperatura inicial,  $T_0$ . A cada interação é gerada uma nova solução, diga-se  $p'$ , na vizinhança da solução corrente  $p$ . Após cada geração de nova vizinhança é avaliada a variação  $\Delta E = f(p') - f(p)$  do valor da função objetivo  $f(\cdot)$ .

---

**Algorithm 2** Pseudocódigo Algoritmo *Simulated Annealing* adaptado.

---

```

begin
  Input:  $p \leftarrow p_0$ ;       $T \leftarrow T_0$ 
  while Enquanto a Temperatura  $T$  for elevada: do
    for interações para o Equilíbrio do
      Gerar uma solução  $p' \in N'(p)$ ;
      Avaliar a geração de energia  $\Delta E = f(p') - f(p)$ ;
      if  $\Delta E < 0$  then
        |  $p \leftarrow p'$ 
      else
        | gerar  $u \in U[0, 1]$ ;
        | if  $u < \exp\left(-\frac{\Delta E}{T}\right)$  then
        | |  $p \leftarrow p'$ 
        | end
      end
    end
    Reduzir a Temperatura  $T$ ;
  end
  Output:  $p_{min}$ 
end

```

---

A função objetivo é tratada como a energia no problema, dessa forma, são gerados novos estados no sistema e aplica-se um conjunto de movimentos, esses movimentos são aceitos ou rejeitados de acordo com o critério de Metrópolis, no qual é baseado o algoritmo SA (WHITE, 1984):

- i) Se  $\Delta E \leq 0$ , há redução de energia, o algoritmo aceita a nova solução, que passa a ser a solução vigente.

- ii) Se  $\Delta E > 0$ , há aumento de energia, então, gera-se um número aleatório  $u$ , que segue uma Uniforme em  $[0, 1]$ , e, se  $u < \exp\left(-\frac{\Delta E}{T}\right)$ , a nova solução é aceita. Como essa solução é pior que a atual, a aceitação é improvável, porém pode ocorrer no início do processo, sob altas temperaturas.

O SA inicia o processo de otimização com uma temperatura muito elevada. Com as repetidas interações, o sistema tende a um equilíbrio térmico naquele estágio de temperatura, e então, a temperatura  $T$  é decrementada por uma razão de resfriamento  $\alpha$ , com  $0 < \alpha < 1$ , dessa forma, a nova temperatura é dada por  $T_n \leftarrow \alpha \cdot T_{n-1}$ , sendo  $T_{n-1}$  a temperatura no estágio anterior. O processo é parado quando a temperatura está suficientemente baixa que nenhuma melhoria útil é esperada (WHITE, 1984).

Nos casos aqui apresentados, a função objetivo que deseja-se minimizar é uma métrica de seleção de modelos, o BIC (Equação 2.84). Os sistemas para os casos de ANN e SVM (individual e combinado) são, respectivamente, o modelo de ANN e o modelo de SVM.

Em ANN, a saída  $p_{min}$  é o vetor  $p = (p_{AR}, p_{ARS}, S, N, LA, AF)$  em que  $p_{AR}$  são os valores Auto-regressivos defasados,  $p_{ARS}$  são os valores Auto-regressivos defasados sazonais,  $S$  é a sazonalidade,  $N$  indica o número de nós na camada intermediária,  $LA$  refere-se ao algoritmo de aprendizagem e  $AF$  a função de ativação, que é o conjunto de parâmetros que minimiza a função objetivo.

As variações consideradas para o algoritmo de aprendizagem foram BPROP, RPROP<sup>+</sup>, RPROP<sup>-</sup>, SAG e SLR. Como variantes da função de ativação foram usadas as funções Logística – LF e Tangente Hiperbólica – THF.

Em SVR, a saída  $p_{min}$  é o vetor  $p = (p_{AR}, p_{ARS}, S, cost, \epsilon, \gamma, degree, coef0, kernel, SV)$  em que  $p_{AR}$ ,  $p_{ARS}$ ,  $S$  são como descritos para ANN,  $cost$  é o custo,  $\epsilon$  é responsável pelo tubo ao redor da função de regressão,  $\gamma$ ,  $degree$  e  $coef0$  são parâmetros das funções Kernel, exceto a linear,  $kernel$  é a função Kernel e  $SV$ , o número de vetores de suporte, que é o conjunto de parâmetros que minimiza a função objetivo.

A variação para o  $costo$  ocorreu no intervalo  $[10^{-5}, 10^2]$ , a do  $\epsilon$  entre  $[10^{-5}, 1]$ , a do  $\gamma$  em  $[10^{-5}, 10^4]$ , do  $degree$  em  $[1, 5]$  e para o  $coef0$  tomou-se valores no intervalo  $[0, 10]$ . Esses valores foram considerados a partir de uma busca na literatura pelos intervalos para otimização dos parâmetros de SVR.

Já as funções Kernel variaram entre linear – LK, polinomial – PK, de base radial – RBFK e sigmóide – SK.

## 2.9 Métricas de Desempenho

Considere  $\hat{u}_t$  o valor estimado para o verdadeiro valor da série no instante  $t$ ,  $u_t$ , e  $n$  o total de observações regressas. Segundo [Firmino, Neto e Ferreira \(2015\)](#), são muitas as medidas existentes na literatura para a avaliação de preditores, a partir das divergências entre  $u_t$  e  $\hat{u}_t$ .

Com o intuito de cobrir o maior número de aspectos que envolvem a avaliação de um preditor, optou-se, neste trabalho, pelo conjunto de métricas formado pelas seguintes medidas: Erro Quadrático Médio (MSE); Erro Médio Absoluto (MAPE); Variância Relativa a Média (ARV); Índice de Desacordo (ID); Coeficiente U de Theil (Theil'U); Predição de Alteração na Direção (WPOCID); Intercepto do ajuste linear entre  $\hat{u}_{t,i}$  e  $u_t$  (Reg\_ Intercept); Coeficiente Angular do ajuste linear entre  $\hat{u}_{t,i}$  e  $u_t$  (Reg\_Slop); Coeficiente de Indeterminação do ajuste linear entre  $\hat{u}_{t,i}$  e  $u_t$  (WR<sup>2</sup>) e a Média Agregada Normalizada (MAN).

O MSE mede a diferença quadrática média entre o valor predito e o valor real da série, sendo uma boa medida para verificar ponderadamente a eficiência e a acurácia de um preditor ([CASELLA; BERGER, 2002](#)). Essa medida é dada por:

$$MSE = \frac{\sum_{t=1}^n (u_t - \hat{u}_t)^2}{n}. \quad (2.73)$$

Por sua vez, o MAPE é uma das medidas de desempenho mais utilizadas, devido às suas vantagens de independência de escala e facilidade de interpretação ([KIM; KIM, 2016](#)). Mede a precisão do preditor em valores relativos. Entretanto, o MAPE é prejudicado por valores iguais ou próximos a zero nos dados observados, produzindo valores infinitos ou indefinidos

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{u_t - \hat{u}_t}{u_t} \right|. \quad (2.74)$$

A métrica Theil compara o desempenho do modelo com um *Random Walk*, do português passeio aleatório, um modelo trivial de predição, que utiliza a observação anterior da série,  $u_{t-1}$ , para prever  $u_t$ . Caso *Theil* > 1, a previsão do modelo é pior que a previsão realizada pelo modelo ingênuo, se *Theil* = 1, o modelo tem desempenho igual ao *Random Walk*, e, se *Theil* < 1, o preditor é melhor que um modelo *Random Walk*, ou seja, quanto mais próximo de zero for esta medida, melhor será a predição.

$$Theil = \frac{\sum_{t=1}^n (u_t - \hat{u}_t)^2}{\sum_{t=1}^n (u_t - \hat{u}_{t-1})^2}. \quad (2.75)$$

Por sua vez, a medida ARV compara as previsões do modelo com a média simples dos valores regressos da série temporal. Sua expressão é dada por

$$ARV = \frac{\sum_{t=1}^n (u_t - \hat{u}_t)^2}{\sum_{t=1}^n (\hat{u}_t - \bar{u}_t)^2}, \quad (2.76)$$

em que  $\bar{u}_t$  é a média da série temporal regressa até o instante  $t$ . Se  $ARV = 1$ , o desempenho do modelo equivale ao de um modelo que sempre retorna a média da série como previsão. Se  $ARV > 1$ , o desempenho do modelo é considerado inferior à previsão realizada utilizando a média da série. E, se  $ARV < 1$ , o modelo utilizado possui desempenho superior a média da série.

O POCID mensura o desempenho do modelo na previsão da tendência da série, ao contrário das métricas anteriores, que se busca a minimização, é desejável que o POCID seja o maior possível. Como alternativa, utilizou-se  $WPOCID = 1 - POCID$ , para as análises se darem no mesmo sentido, o de minimizar essas métricas. Essa medida é definida por

$$POCID = \frac{\sum_{t=1}^n D_t}{n}, \quad (2.77)$$

em que  $D_t = 1$  se  $(\hat{u}_t - \hat{u}_{t-1})(u_t - u_{t-1}) > 0$  (indicando que o modelo acertou a previsão de tendência da série entre os instantes  $t - 1$  e  $t$ ) e  $D_t = 0$  se  $(\hat{u}_t - \hat{u}_{t-1})(u_t - u_{t-1}) < 0$  (indica que o modelo errou a previsão de tendência).

Considerou-se também a métrica ID – Índice de Desacordo, medida complementar ao Índice de Concordância – IC (WILLMOTT, 1982), usado para comparar diversos modelos de um mesmo fenômeno, independente da unidade de medida, uma vez que seu valor é dado no intervalo  $[0, 1]$ . Essa medida indica a precisão do valor predito em relação ao observado e se  $IC = 1$  a concordância é perfeita e  $IC = 0$  indica total discordância entre observado e previsto. No caso do ID ocorre o contrário: quanto mais próximo de zero estiver essa medida, melhor é o modelo. Sua expressão é definida por

$$ID = 1 - IC \quad (2.78)$$

$$= 1 - \left\{ 1 - \frac{\sum_{t=1}^n (\hat{u}_t - u_t)^2}{\sum_{t=1}^n (|\hat{u}_t - \bar{u}_t| + |u_t - \bar{z}_t|)^2} \right\} \quad (2.79)$$

$$= \frac{\sum_{t=1}^n (\hat{u}_t - u_t)^2}{\sum_{t=1}^n (|\hat{u}_t - \bar{u}_t| + |u_t - \bar{z}_t|)^2}. \quad (2.80)$$

Foram ainda consideradas as métricas  $RegWR2 = 1 - R^2$ , bem como Reg\_Intercept e Reg\_Slope, relacionados ao modelo de ajuste linear entre  $u_t$  e  $\hat{u}_t$ , através do métodos

de mínimos quadrados. Considere uma equação geral da forma  $u_t = a + b \cdot \hat{u}_t$ . Assim,  $a$  e  $b$  são coeficientes que representam os erros aditivo (Reg\_ Intercept) e multiplicativo (Reg\_ Slope) cometidos pelo modelo quando este gera a previsão  $\hat{u}_t$  para  $u_t$ . Nesse caso, há um erro constante  $a$  independente da previsão e um erro proporcional  $b$  relacionado à previsão. Por sua vez,  $R^2$ , o coeficiente de determinação, reflete o desempenho do modelo quanto à captura da variabilidade da série temporal observada (NETO et al., 2020). O  $R^2$  é definido como

$$R^2 = \frac{\sum_{t=1}^n (u_t - \hat{u}_t)^2}{\sum_{t=1}^n (u_t - \bar{u})^2}, \quad (2.81)$$

em que  $\bar{u}$  representa a média de toda a série temporal observada. Logo, um preditor ideal teria como medidas  $R^2 = 1$ ,  $a = 0$  e  $b = 1$ , obtendo assim,  $u_t = 0 + 1 \cdot \hat{u}_t$ .

Para agregar a análise dessas medidas de desempenho será usada ainda uma Média Agregada Normalizada – MAN que agrega todas as medidas acima, normalizadas no intervalo  $[0, 1]$ , e calcula a média simples destas, gerando um resultado único para cada modelo, também com valores entre 0 e 1. Quanto mais próxima de 0 estiver essa métrica, melhor será o modelo avaliado e, do contrário, quanto mais próxima de 1 estiver, pior será o desempenho do modelo. Define-se essa métrica através da expressão

$$MAN = \frac{\sum_{i=1}^m (NMétrica_i)}{m}, \quad (2.82)$$

em que  $NMétrica_i$  são as métricas normalizadas consideradas para o cálculo e  $m = 8$  refere-se à quantidade de métricas (MSE, MAPE, ARV, ID, Theil'U, WPOCID, |Reg\_ Intercept| e ||Reg\_ Slope|-1|). A normalização para o intervalo  $[0, 1]$  se dá através do cálculo

$$NMétrica_i = \frac{Métrica_i - \min_i}{\max_i - \min_i}, \quad (2.83)$$

em que  $\min_i$  e  $\max_i$  são respectivamente os valores mínimo e máximo observados para  $Métrica_i$  ao longo dos modelos ajustados à série temporal sob estudo. Em todas as métricas consideradas o objetivo é a minimização das equações que as expressam, de forma geral, o melhor modelo será o que minimizar a Equação 2.82.

## 2.10 Critério de Informação Bayesiano

Na modelagem estatístico-computacional é comum surgir o desafio de escolher um modelo mais adequado dentre os possíveis candidatos. Para simplificar essa tarefa faz-se aqui o uso de um critério de seleção (NEATH; CAVANAUGH, 2012). Para a seleção dos

modelos individuais e combinados utilizou-se o Critério de Informação Bayesiano (BIC), proposto por Schwarz (1978), dado por:

$$BIC = -2 \ln f(\mathbf{e}|\hat{\theta}) + k \ln(n), \quad (2.84)$$

em que  $f(\mathbf{e}|\hat{\theta})$  é o valor da função de verossimilhança diante dos resíduos observados ( $\mathbf{e}$ ) e das estimativas dos parâmetros desta função ( $\hat{\theta}$ ),  $k$  é a quantidade de parâmetros de  $f(\cdot)$  e  $n$  é a quantidade de observações da amostra. O modelo que minimiza a Equação 2.84 será o escolhido.

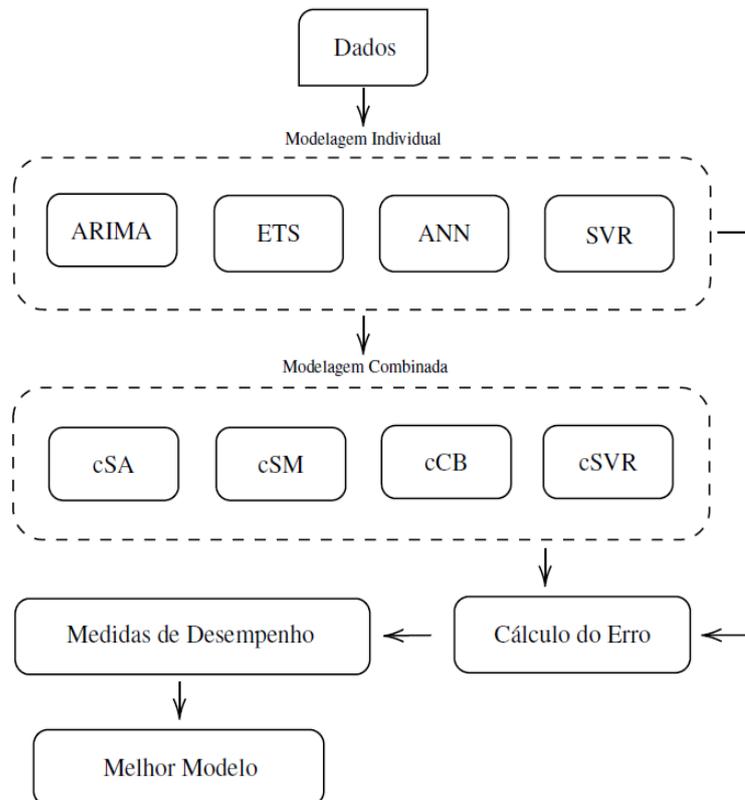
O BIC é preferido em detrimento de outros critérios de informação devido a sua tendência em escolher modelos mais parcimoniosos, sendo uma ferramenta científica muito eficaz na seleção de modelos (NEATH; CAVANAUGH, 2012). Ele foi adotado tanto como função objetivo dos métodos de otimização, na busca por preditores baseados em ANN e SVR, quanto para o ajuste dos preditores baseados em cópulas (cCB).

## 3 Experimentação

### 3.1 Arquitetura do estudo

Os dados utilizados neste estudo são compostos por doze séries temporais sobre variáveis ligadas ao Desenvolvimento Sustentável, que serão analisadas sob a óptica da teoria de séries temporais, através do procedimento descrito nas etapas a seguir, resumido pela Figura 10, usado para modelar as variáveis em questão, obtendo modelos preditivos individuais e combinados. Todas as análises, gráficos e tabelas foram geradas no *software* gratuito R (COMPUTING, 2015), através da interface RStudio.

Figura 10 – Arquitetura proposta para modelagem individual e combinada.



Fonte: Autor (2020).

O percurso para a realização da experimentação adotado nesse trabalho encontra-se descrito nos sete tópicos a seguir:

1. Destinou-se 80% dos dados para a etapa de treinamento, na qual os preditores são construídos. Os 20% restantes foram destinados para o teste dos modelos, diante de dados desconhecidos. Todas as séries de treinamento foram normalizadas no intervalo fechado  $[0, 4; 0, 6]$ . A adoção desse intervalo vem da necessidade das entradas para modelagem via ANN estarem entre 0 e 1, e, caso haja extrapolação do intervalo  $[0, 4; 0, 6]$ , os valores ainda tenderão a permanecer no intervalo  $[0, 0; 1, 0]$ , na etapa de teste.
2. Os modelos individuais ARIMA e ETS foram implementados através do pacote `forecast` (HYNDMAN; KHANDAKAR, 2008), respectivamente, pelas funções `auto.arima` e `ets`. Os melhores modelos foram selecionados através do BIC.
3. Por sua vez, outros modelos de Machine Learning para modelagem individual, ANN e SVR, foram implementados via pacotes `neuralnet` (FRITSCH; GUENTHER; WRIGHT, 2019) e `e1071` (MEYER et al., 2019), respectivamente, através das funções `neuralnet` e `svm`. Tais modelos foram otimizados via pacote `GenSA` (XIANG et al., 2013), que busca a minimização do BIC. Para variações do AL na ANN foram adotados: BPROP, RPROP+, RPROP-, SAG e SLR e para a FA tomou-se FL, FS e FTH. No caso do SVR, o tipo adotado foi *eps - regression*, utilizado para casos de regressão via SVR e as variações para função Kernel foram: Linear, Polynomial, Sigmoid e Radial Basis.
4. As combinações via Média Simples (cSA) e Mediana Simples (cSM) levaram em consideração as previsões dos modelos individuais, calculando-se, respectivamente, a média e a mediana destes.
5. Para a previsão combinada paramétrica foram consideradas duas metodologias: Cópula (cCB) e SVR (cSVR). cCB leva em consideração a função de distribuição acumulada dos erros dos modelos individuais e cria uma distribuição conjunta. Para inferência dos parâmetros da cópula, utilizou-se o método IFM, o qual estabelece que os parâmetros das marginais sejam inferidos separadamente para cada modelo individual. As Famílias de cópulas adotadas foram as CA, com as cópulas Clayton, Frank, Gumbel e Joe e as CE, com a cópula Normal. Para as marginais, optou-se pelas distribuições Normal, Normal-assimétrica, Laplace e Cauchy. Para esse caso, foi utilizado o pacote `copula` (HOFERT; MÄCHLER, 2011).
6. A combinação via cSVR ocorre de maneira semelhante à modelagem individual, porém agora, as entradas serão as previsões dos modelos individuais. Os modelos combinados via cSVR foram otimizados de forma similar às SVR's individuais. A variação para o *custo* ocorreu no intervalo  $[10^{-5}, 10^2]$ , a do  $\epsilon$  entre  $[10^{-5}, 1]$ , a do

$\gamma$  em  $[10^{-5}, 10^4]$ , do *degree* em  $[1, 5]$  e para o *coef0* tomou-se valores no intervalo  $[0, 10]$ . Para as funções kernel considerou-se LK, PK, RBFK e SK.

- Finalmente, os modelos individuais e combinados foram comparados segundo as métricas de desempenho MSE, MAPE, ARV, ID, Theil'U, WPOCID, Reg\_Intercept, Reg\_Slope e MAN.

## 3.2 Conjunto de Dados

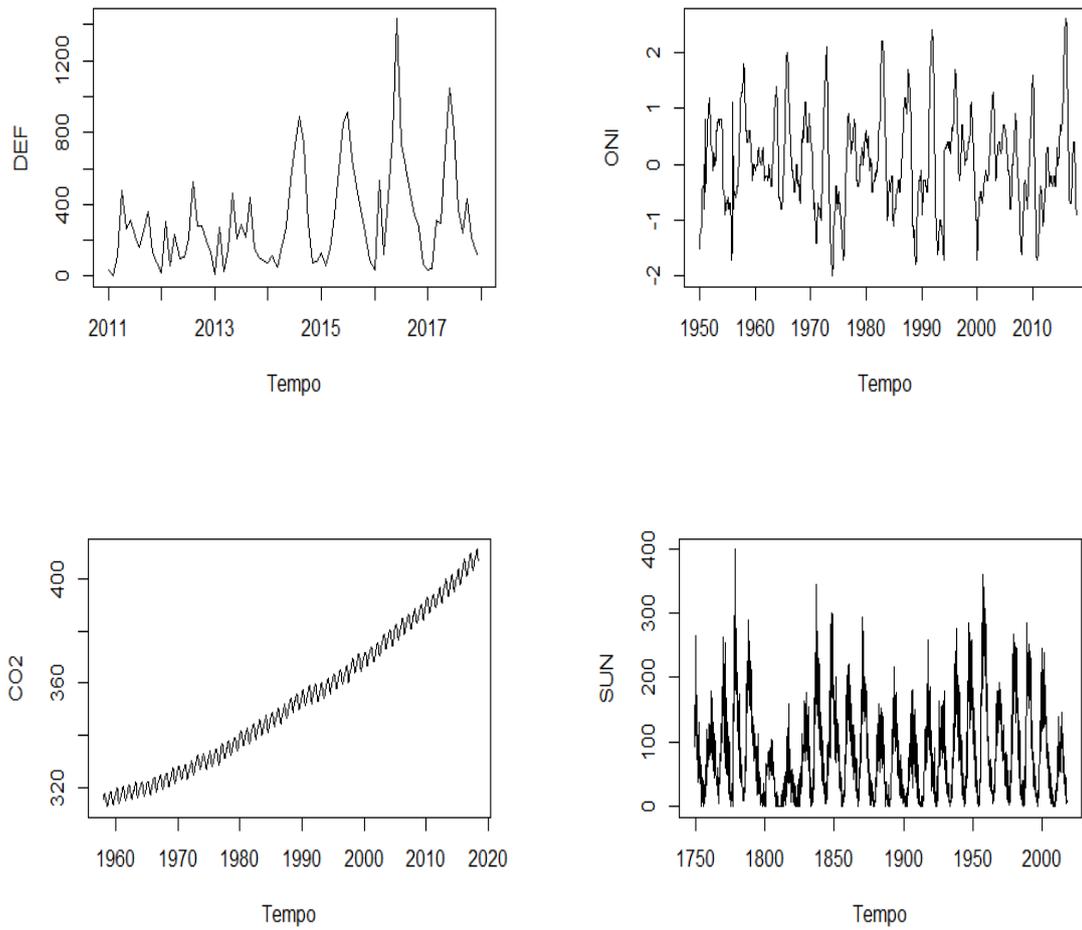
As séries temporais utilizadas neste trabalho versam sobre os três principais pilares do Desenvolvimento Sustentável: Ambiental, Econômico e Social. Tratam-se de dados reais que serão uma fonte para análise e comparação acerca da qualidade dos preditores gerados. As séries do Eixo Ambiental são: Desmatamento da Amazônia Brasileira (DEF) (DETER, 2018), Índice de Oscilação Sul do El Niño (ONI) (NOOA, 2018a), Concentração de Dióxido de Carbono na Atmosfera (CO<sub>2</sub>) (NOOA, 2018b), Índice de Manchas Solares (SUN)(TONG, 1990). No Eixo Econômico tem-se: Produção Brasileira de Biodiesel (BBP) (ANP, 2018), Índice de Preços ao Consumidor no Brasil (IPCA)(IBGE, 2018), Preço das Ações da Petrobras (PSP) (ATREDINGVIEW, 2018) e Índice S&P500 (SP500) (YAHOOFINANCE, 2018). Por fim, no Eixo Social: Mortes em Veículos Motorizados do Ceará - Brasil (MVF) (DETRAN, 2018), Número de Casos de Dengue no Rio de Janeiro - Brasil (DEN) (SMS-RJ, 2018), Consumo de Energia no Nordeste Brasileiro (ENE) (IPEA, 2018) e Número de Casos de Caxumba em Nova York (MUC) (DATAMARKET, 2018). A Tabela 2 apresenta um resumo das características das séries e nas Figuras 11, 12 e 13 é possível observar as séries temporais graficamente.

Tabela 2 – Resumo das Características das Séries Temporais em Estudo

Série	Tot. observações	Periodicidade	Conjunto de treinamento	Conjunto de teste
DEF	84	mensal	67	17
ONI	814	mensal	652	162
CO <sub>2</sub>	726	mensal	580	145
SUN	3240	mensal	2592	648
BBP	144	mensal	115	29
IPCA	226	mensal	181	45
SPS	216	mensal	173	43
SP500	1006	diária	805	201
MVF	84	mensal	67	17
DEN	161	mensal	129	32
ENE	476	mensal	381	95
MUC	534	mensal	427	107

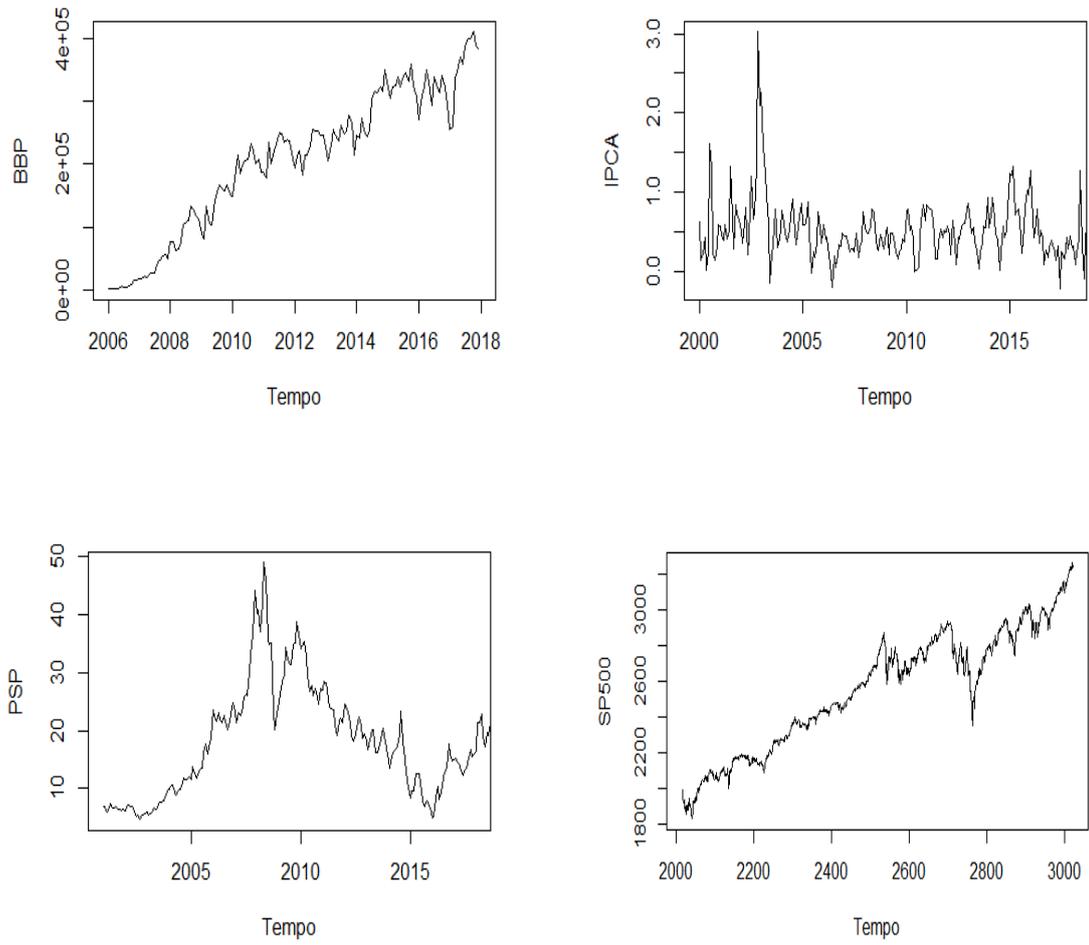
Fonte: Autor (2020).

Figura 11 – Séries Temporais Ambientais.



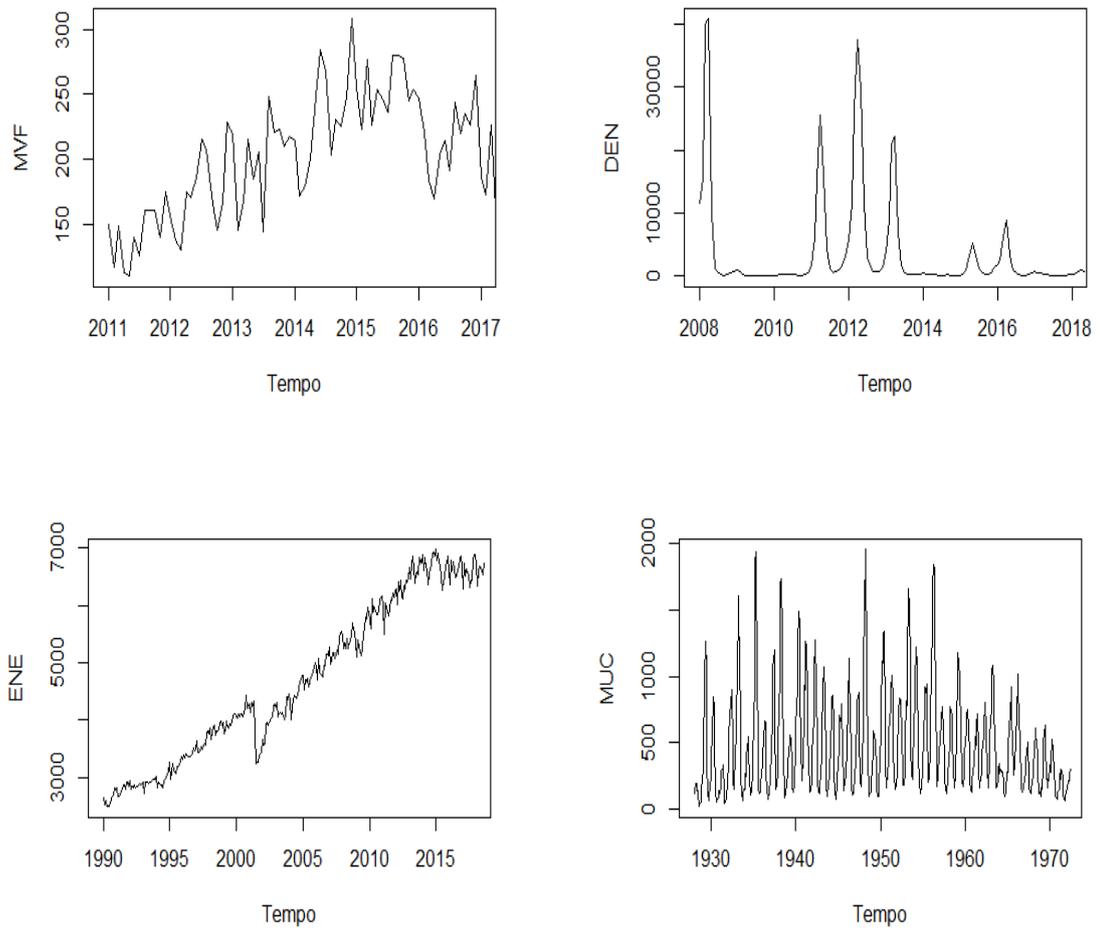
Fonte: Autor (2020).

Figura 12 – Séries Temporais Econômicas.



Fonte: Autor (2020).

Figura 13 – Séries Temporais Sociais.



Fonte: Autor (2020).

## 4 Resultados

Para realização do estudo foram consideradas doze séries temporais ligadas a variáveis do Desenvolvimento Sustentável, nas áreas ambiental, econômica e social, com diferentes tamanhos, com o intuito de analisar o comportamento dos modelos propostos sob perspectivas distintas.

Foram gerados um total de 96 modelos, sendo 8 para cada série, 4 individuais e 4 combinadores. A seguir serão apresentados a estrutura e o desempenho de cada modelo em cada série. Quanto à configuração do computador adotado para a experimentação, tratou-se de um notebook, sistema operacional Windows 10 Home (64 bits), processador Intel i7, com 2.6GHz e 8GB de memória RAM.

A Tabela 3 mostra a configuração utilizada na etapa de modelagem das séries.

Tabela 3 – Parâmetros de ajuste dos modelos.

Parâmetro	Valor	modelo
Proporção de treinamento	0.8	todos
Proporção AR e ARS	0.05	ANN, SVR
ANN.threshold	0.001	ANN
ANN.stepmax	6000000	ANN
Número de modelos	5000	ARIMA, ETS
GSA.max.call	10000	ANN, SVR, cSVR
GSA.max.time	600	ANN, SVR, cSVR
GSA.maxit	1000000	ANN, SVR, cSVR
GSA.temperature	10000	ANN, SVR, cSVR
GSA.nb.stop.improvement	20	ANN, SVR, cSVR

Fonte: Autor (2020).

Os modelos combinados cSA e cSM não necessitam de etapa de modelagem; dessa forma, não foram considerados para a construção da Tabela anterior. Ademais, na apresentação que se segue dos modelos e suas respectivas descrições, esses não aparecem por não apresentarem parâmetros passíveis de interpretação.

## 4.1 Séries Temporais Ambientais

### 4.1.1 Desmatamento da Amazônia Brasileira

O desmatamento da Amazônia brasileira se intensificou a partir de 1991, variando em consonância com as mudanças econômicas. São muitas as razões que elevaram os índices de desmatamento, incluindo decisões governamentais. A ocorrência do desmatamento na floresta leva a perda de serviços ambientais, uma vez que sua área devastada é substituída por atividades pouco sustentáveis (FEARNSIDE, 2006).

As consequências da devastação da Amazônia incluem perda de produtividade, mudanças no regime hidrológico, perda da biodiversidade e aumento da emissão de gases do efeito estufa. Dessa forma, é essencial o controle do desmatamento para a manutenção da Amazônia, entretanto não existe uma solução única, são necessárias um conjunto de medidas que incluem a prevenção do desflorestamento, suspensão das ações governamentais que o fomentam e oferta de alternativas àqueles que dependem da agricultura de subsistência (FEARNSIDE, 2005; FEARNSIDE, 2017).

Na Tabela 4 encontram-se descritas as arquiteturas dos modelos individuais e combinados para a série DEF. O modelo ARIMA não apresentou termos autorregressivos ou de médias móveis, indicando apenas a necessidade de uma diferenciação para estabilizar os valores observados. O modelo ETS indicou um erro multiplicativo e nenhuma sazonalidade ou tendência. Logo, ambos os modelos se assemelham ao modelo de *random walk*.

Por sua vez, os demais modelos de *Machine Learning*, ANN e SVR, apresentaram uma entrada autorregressiva, o que significa que foi utilizado apenas o valor anterior,  $u_{t-1}$ , para realizar a previsão da série no instante  $t$  ( $u_t$ ).

Na combinação via cópula as distribuições marginais dos erros dos quatro preditores individuais foram Normais ( $N$ ) e Normais assimétricas ( $Norms$ ), com médias e desvios padrão variados, e a distribuição conjunta foi a cópula de Frank.

Tabela 4 – Descrição dos modelos individuais e combinados para a série DEF.

Formalismo	Modelo	Descrição
ANN	ANN(1, 1, 1)	$P_{AR} = 1, P_{ARS} = 0, LA = SAG, FA = LF$
ARIMA	ARIMA(0, 1, 0)	-
ETS	ETS(M, N, N)	$\alpha = 0.905$
SVR	-	$P_{AR} = 1, P_{ARS} = 0, cost = 89.6793, \epsilon = 0.9835,$ $kernel = LK, SV = 11$
cCB	-	$E_{ANN} \sim NormS(média = -1.6 \times 10^{-3}, sd = 0.0253,$ $\xi = 1.7012),$ $E_{ARIMA} \sim N(1.6 \times 10^{-3}, sd = 0.0277),$ $E_{ETS} \sim N(1.7 \times 10^{-3}, sd = 0.0275),$ $E_{SVR} \sim NormS(média = -6.1 \times 10^{-3}, sd = 0.0253,$ $\xi = 1.9001),$ $Cópula = Frank(\alpha = 20.0343)$
cSVR	-	$cost = 73.6556, \epsilon = 0.9914, kernel = PK, SV = 9$ $\gamma = 6834.1956, degree = 1.3016, coef0 = 4.1416$

Fonte: Autor (2020).

A combinação cSVR considerou como entradas as previsões dos quatro modelos individuais e obteve um modelo com Kernel Polinomial e 9 vetores de suporte. Na Tabela 5 são mostradas as métricas de desempenho dos modelos individuais e combinadores para efeito de comparação. Dentre os preditores individuais, o ARIMA destacou-se sobre os demais, já para os modelos combinados, a cSA apresentou melhores resultados. Tal resultado pode estar atrelado à dificuldade dos modelos individuais se distanciarem de um *random walk*.

Tabela 5 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série DEF (fase de teste). Os melhores valores encontram-se em negrito.

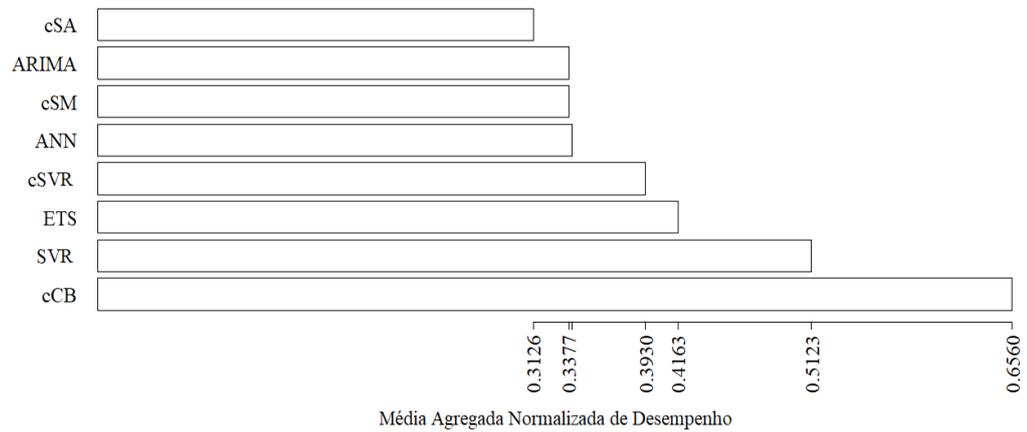
Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	<b>38651</b>	47577	51644	40026	41361	41702	70809	40878
MAPE	0.872	0.661	0.721	1.055	0.820	0.790	<b>0.626</b>	0.939
ARV	2.194	<b>0.904</b>	1.048	2.500	1.345	1.353	1.807	2.007
ID	0.295	<b>0.239</b>	0.274	0.316	0.263	0.264	0.429	0.300
Theil	<b>0.831</b>	1.000	1.063	0.861	0.882	0.891	1.431	0.876
WPOCID	0.375	0.375	<b>0.312</b>	0.375	0.375	0.375	0.375	0.375
Reg_Intercept	-37.175	95.755	99.255	-92.842	36.148	43.284	<b>9.002</b>	-22.100
Reg_Slope	1.080	0.679	0.659	1.150	0.838	0.830	1.703	<b>0.974</b>
Reg_WR2	<b>0.523</b>	0.523	0.560	0.523	0.533	0.540	0.535	0.541

Fonte: Autor (2020).

A Figura 14 ilustra a Média Agregada Normalizada de desempenho dos preditores.

Nesta série, a cCB obteve o pior desempenho geral entre os 8 preditores, seguida de SVR, enquanto a cSA foi o melhor modelo.

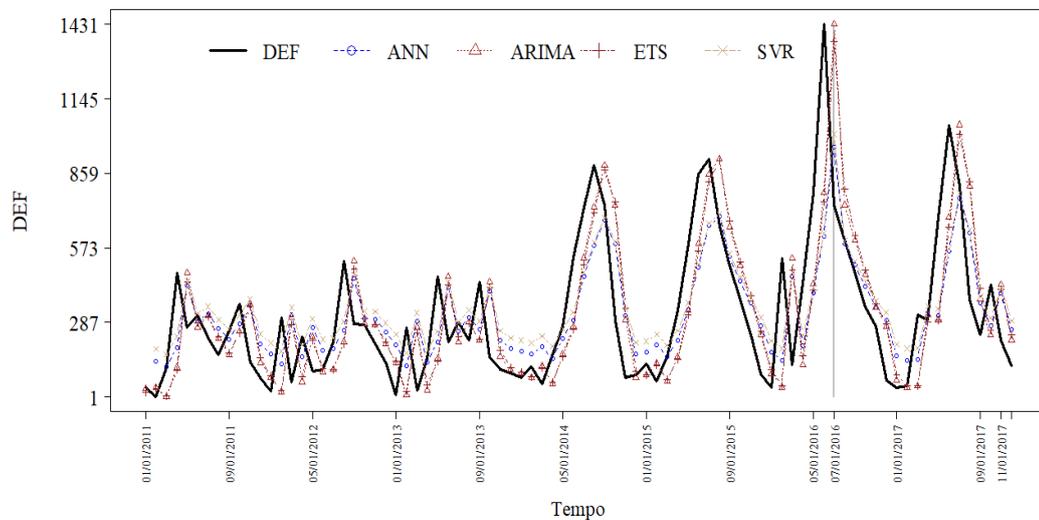
Figura 14 – Média Agregada de Performance dos modelos individuais e combinados para a série DEF.



Fonte: Autor (2020).

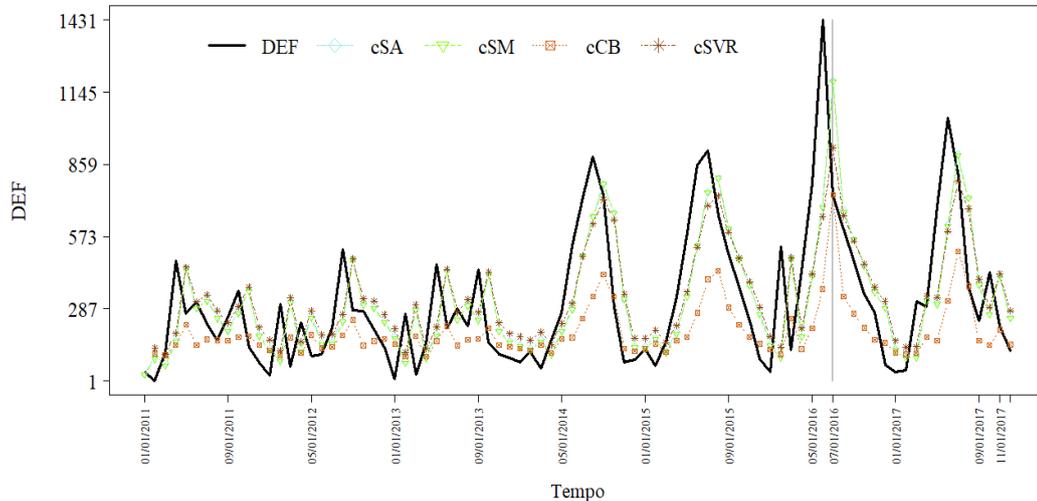
A Figura 15 ilustra a série DEF e as previsões modelos preditivos individuais. As previsões dos modelos combinados são mostradas na Figura 16.

Figura 15 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série DEF (Fase de teste).



Fonte: Autor (2020).

Figura 16 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série DEF (Fase de teste).



Fonte: Autor (2020).

#### 4.1.2 Índice Oceânico de *Niño*

O ciclo *El Niño*-Oscilação do Sul (ENSO) é um fenômeno cíclico da atmosfera oceânica originado no Pacífico tropical e possui duas fases opostas: *El Niño* e *La Niña*. O *El Niño* se caracteriza pelo aquecimento das temperaturas na superfície do mar no Pacífico equatorial central e leste leste, enquanto o *La Niña* é responsável pelo resfriamento dessas águas. O ENSO pode gerar anomalias nos padrões de precipitação e temperatura em escala mundial (GUTIERREZ, 2017).

O Índice Oceânico de *Niño* – ONI é determinado pela média consecutiva de três meses de anomalias de temperatura na superfície do mar que configura o *El Niño* ou *La Niña*, as anomalias devem exceder  $+0,5^{\circ}\text{C}$  ou  $-0,5^{\circ}\text{C}$  por pelo menos cinco meses consecutivos. Este indicador pode ser usado para analisar os impactos das anomalias do ENSO em diversos setores e a previsão antecipada desse fenômeno é de extrema importância devido as suas implicações para a sociedade, economia e ecossistema (SUN et al., 2018; MENG et al., 2020).

A Tabela 6 mostra as arquiteturas dos modelos individuais e combinados para a série ONI. O modelo ARIMA não apresentou termos autorregressivos e nem diferenciações, mas um modelo de médias móveis com 19 termos. O modelo ETS apresentou um erro do tipo aditivo, tendência aditiva amortecida e nenhuma sazonalidade. O modelo ANN foi configurado com 7 nós na camada intermediária, já o modelo SVR considerou duas entradas autorregressivas e 19 vetores de suporte.

Tabela 6 – Descrição dos modelos individuais e combinados para a série ONI.

Formalismo	Modelo	Descrição
ANN	ANN(8, 7, 1)	$P_{AR} = 8, P_{ARS} = 0, LA = SLR, FA = THF$
ARIMA	ARIMA(0, 0, 19)	-
ETS	ETS(A, Ad, N)	$\alpha = 0.921, \beta = 0.283, \phi = 0.876$
SVR	-	$P_{AR} = 2, P_{ARS} = 0, cost = 77.4876, \epsilon = 0.5445,$ $kernel = LK, SV = 19$
cCB	-	$E_{ANN} \sim N(-1 \times 10^{-4}, sd = 6.3 \times 10^{-3}),$ $E_{ARIMA} \sim N(7.1 \times 10^{-3}, sd = 7.2 \times 10^{-3}),$ $E_{ETS} \sim N(0, sd = 7.5 \times 10^{-3}),$ $E_{SVR} \sim N(8 \times 10^{-4}, sd = 7.9 \times 10^{-3}),$ $Cópula = Normal (\rho_1 = 0.3455, \rho_2 = 0.7569,$ $\rho_3 = 0.7826, \rho_4 = 0.598, \rho_5 = 0.3502, \rho_6 = 0.7143)$
cSVR	-	$cost = 58.5034, \epsilon = 0.83, kernel = PK, SV = 48$ $\gamma = 8947.9495, degree = 3.3948, coef0 = 4.4552$

Fonte: Autor (2020).

A Tabela 7 mostra as medidas de desempenho dos modelos, com destaque para o modelo ANN que apresentou melhores resultados dentre os modelos individuais e cCB dentre os combinados, que também foi o melhor de forma geral, mesmo não apresentando melhor *WPOCID*, ou seja, errando mais a previsão da tendência da série que o ETS. A Figura 17 resume o desempenho dos modelos, indicando que o modelo cSVR foi o pior entre todos.

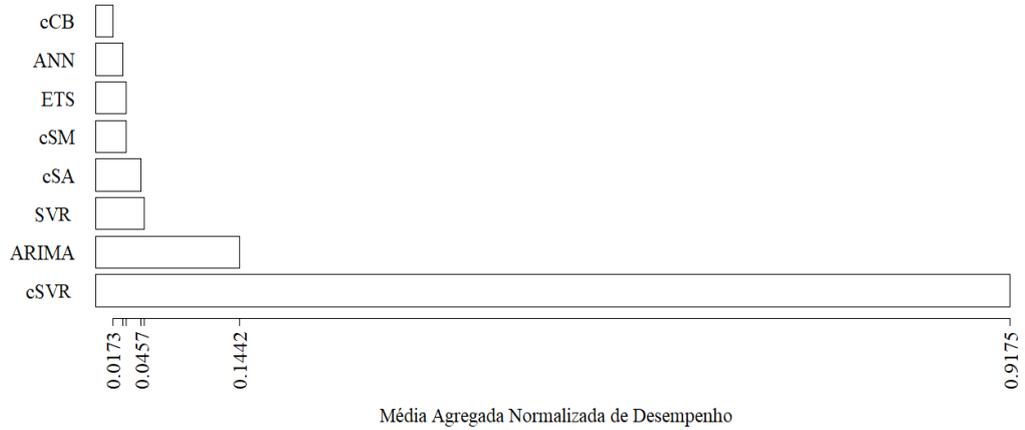
Tabela 7 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série ONI (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	0.026	0.041	0.033	0.038	0.021	0.023	<b>0.017</b>	6.301
MAPE	0.336	0.476	0.404	0.401	0.331	0.351	<b>0.279</b>	2.155
ARV	0.038	0.051	0.040	0.063	0.030	0.032	<b>0.024</b>	0.679
ID	0.009	0.014	0.011	0.014	0.007	0.008	<b>0.006</b>	0.448
Theil	0.555	0.870	0.703	0.812	0.453	0.489	<b>0.361</b>	135.553
WPOCID	0.278	0.265	<b>0.247</b>	0.278	0.272	0.265	0.278	0.481
Reg_Intercept	<b>-0.001</b>	0.148	-0.006	0.004	0.038	0.019	-0.001	-0.039
Reg_Slope	1.020	0.961	0.924	1.063	0.998	<b>1.000</b>	1.015	0.198
Reg_WR2	0.034	<b>0.021</b>	0.038	0.047	0.026	0.030	0.022	0.511

Fonte: Autor (2020).

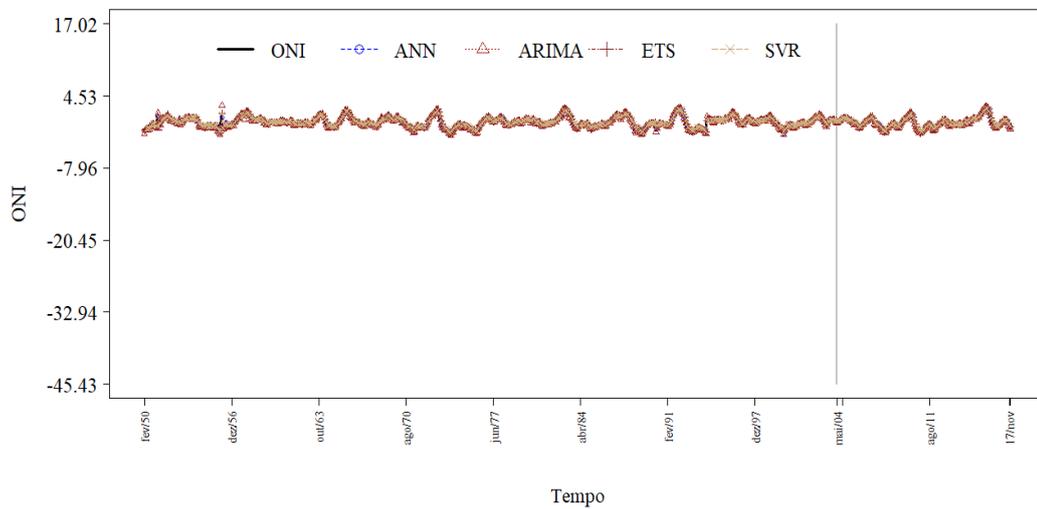
Nas Figuras 18 e 19 encontram-se as representações gráficas das previsões dos modelos individuais e combinados, respectivamente, para a série ONI.

Figura 17 – Média Agregada de Performance dos modelos individuais e combinados para a série ONI.



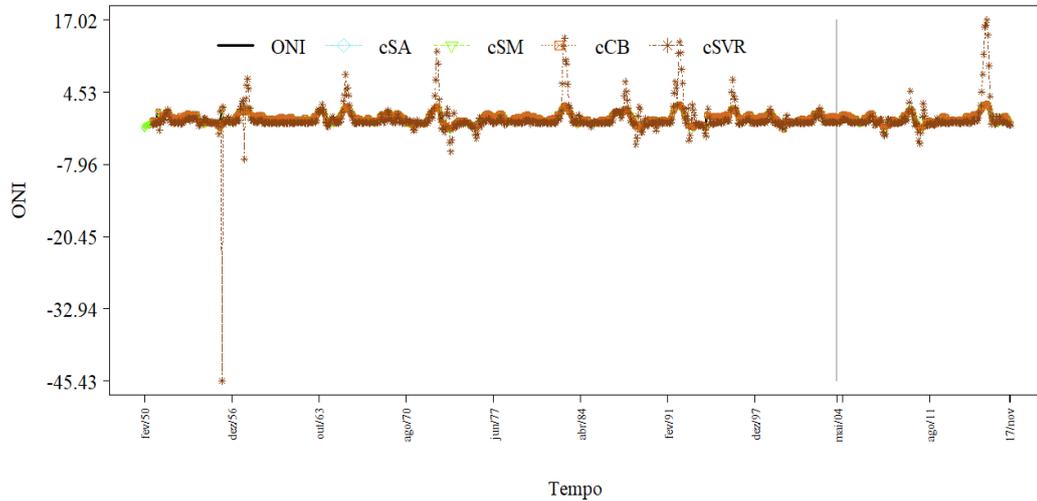
Fonte: Autor (2020).

Figura 18 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série ONI (Fase de teste).



Fonte: Autor (2020).

Figura 19 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série ONI (Fase de teste).



Fonte: Autor (2020).

### 4.1.3 Concentração de Dióxido de Carbono na Atmosfera

As mudanças climáticas se tornaram motivo de preocupação essencial para a sociedade. Eventos extremos têm ocorrido devido à alta emissão de gases do efeito estufa na atmosfera. O gás carbônico, ou dióxido de carbono –  $\text{CO}_2$  é um dos principais gases do efeito estufa e sua emissão tem aumentado continuamente nos últimos anos, mesmo com inúmeras políticas públicas para redução. Com o crescimento da população mundial é natural que os níveis de gases do efeito estufa sejam elevados, uma vez que há aumento de demanda de energia entre outros. A principal fonte de emissão de  $\text{CO}_2$  é a combustão de combustíveis fósseis, como petróleo e gás natural (LI et al., 2018).

A previsão da emissão de  $\text{CO}_2$  é uma preocupação mundial e uma ferramenta de conscientização pública na tentativa de solucionar os problemas climáticos, portanto é de suma importância uma previsão confiável de suas emissões futuras (NYONI; BONGA, 2019).

Na literatura é possível encontrar modelos ARIMA (LI et al., 2018) e de ANN (NGUYEN; HALEM, 2018) para previsão de  $\text{CO}_2$ . Na Tabela 8 estão descritos os modelos individuais e combinados para a série  $\text{CO}_2$ . O modelo ANN utilizou 14 entradas autorregressivas e apenas um nó na camada intermediária, sendo o que mais acertou a previsão de tendência da série ( $WPOCID = 0.035$ ). O ARIMA, por sua vez, utilizou dois termos autorregressivos, um de média móvel e uma diferenciação para estabilizar a série. No modelo ETS, adota-se um erro aditivo e nenhuma sazonalidade ou tendência, se

aproximando de um *random walk*.

Os modelos SVR e cSVR utilizaram a mesma função kernel, porém com parâmetros distintos. SVR utilizou apenas 5 vetores de suporte, enquanto cSVR usou 12. O modelo de cCB ajustou as distribuições marginais dos erros dos modelos como Normais e obteve uma distribuição conjunta também Normal.

Tabela 8 – Descrição dos modelos individuais e combinados para a série CO<sub>2</sub>.

Formalismo	Modelo	Descrição
ANN	ANN(14, 1, 1)	$P_{AR} = 13, P_{ARS} = 1, S = 24, LA = SLR, FA = LF$
ARIMA	ARIMA(2, 1, 1)	-
ETS	ETS(A, N, N)	$\alpha = 1.0$
SVR	-	$P_{AR} = 13, P_{ARS} = 5, S = 24, cost = 20.4725, \epsilon = 0.0877,$ $kernel = PK, SV = 5, degree = 1.7707, coef0 = 0.4813$ $\gamma = 1589.2588$
cCB	-	$E_{ANN} \sim N(0, sd = 1.4 \times 10^{-3}),$ $E_{ARIMA} \sim N(4 \times 10^{-4}, sd = 1.7 \times 10^{-3}),$ $E_{ETS} \sim N(3 \times 10^{-4}, sd = 3.3 \times 10^{-3}),$ $E_{SVR} \sim N(4 \times 10^{-4}, sd = 1.8 \times 10^{-3}),$ $Cópula = Normal(\rho_1 = 0.253, \rho_2 = 0.2177,$ $\rho_3 = 0.7019, \rho_4 = 0.5343, \rho_5 = 0.1953, \rho_6 = 0.2994)$
cSVR	-	$cost = 31.9193, \epsilon = 0.0726, kernel = PK, SV = 12$ $\gamma = 808.466, degree = 3.8615, coef0 = 9.7652$

Fonte: Autor (2020).

A Tabela 9 mostra o desempenho dos modelos individuais e combinados para a série CO<sub>2</sub>, da qual pode-se observar o desempenho superior do SVR, que foi o melhor modelo entre todos. Com destaque para o pequeno número de erros cometidos pelo modelo (ver *MSE* e *MAPE*).

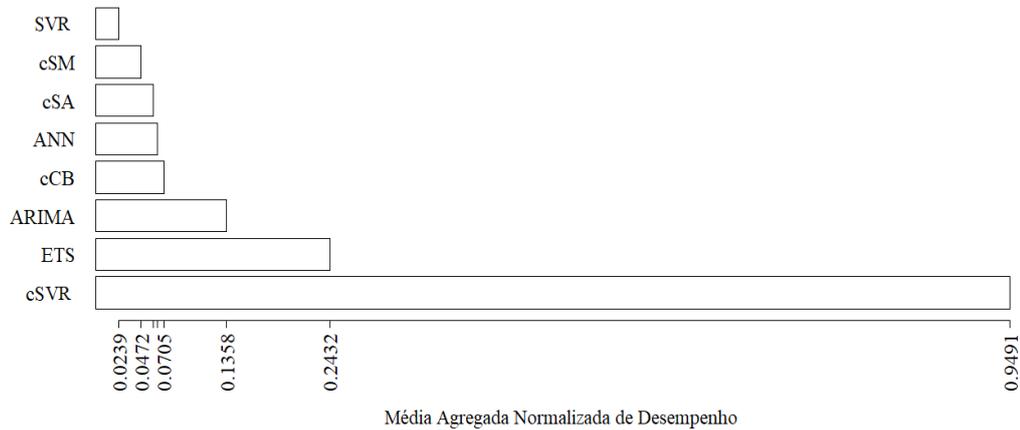
Tabela 9 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série CO<sub>2</sub> (fase de teste). Os melhores valores encontram-se em negrito

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	1.830	0.605	1.764	<b>0.539</b>	0.650	0.610	0.845	171.720
MAPE	0.003	0.002	0.003	<b>0.001</b>	0.002	0.002	0.002	0.026
ARV	0.037	0.009	0.026	<b>0.009</b>	0.011	0.010	0.015	0.380
ID	0.008	0.002	0.007	<b>0.002</b>	0.003	0.002	0.003	0.197
Theil	1.054	0.348	1.000	<b>0.310</b>	0.373	0.351	0.487	98.981
WPOCID	<b>0.035</b>	0.167	0.167	0.049	0.076	0.069	0.083	0.125
Reg_Intercept	-37.8	<b>1.1</b>	5.0	-9.9	-11.7	-10.2	-22.0	185.6
Reg_Slope	1.099	<b>0.998</b>	0.988	1.026	1.031	1.027	1.057	0.516
Reg_WR2	0.003	0.008	0.025	0.003	0.005	0.004	<b>0.003</b>	0.022

Fonte: Autor (2020).

Nota-se, entre os modelos individuais, que o ANN teve pior resultado e dentre os combinados o cSM foi o melhor para a série. Por sua vez, cSVR foi o pior modelo, de acordo com a métrica agregada (ver Figura 20).

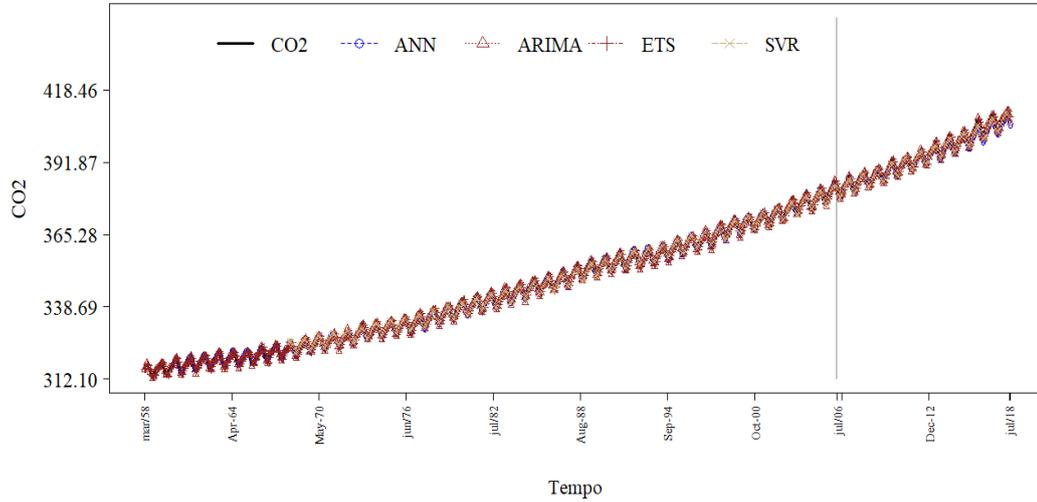
Figura 20 – Média Agregada de Performance dos modelos individuais e combinados para a série CO<sub>2</sub>.



Fonte: Autor (2020).

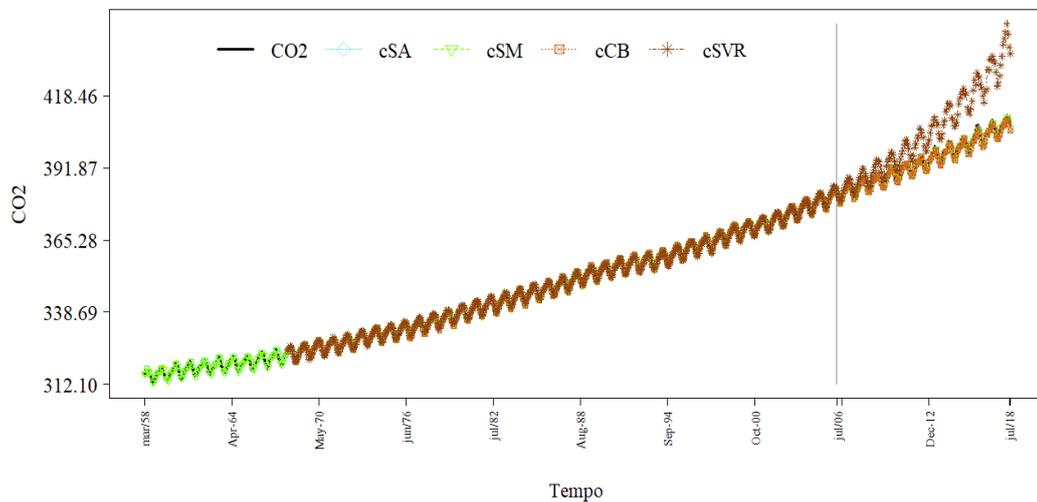
A representação gráfica das previsões dos modelos individuais e combinados para a série CO<sub>2</sub> pode ser observada nas Figuras 21 e 22.

Figura 21 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série CO<sub>2</sub> (Fase de teste).



Fonte: Autor (2020).

Figura 22 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série CO<sub>2</sub> (Fase de teste).



Fonte: Autor (2020).

#### 4.1.4 Índice de Manchas Solares

O índice de manchas solares é o principal indicador do nível de atividade solar, fator este que pode influenciar a saúde humana e o meio ambiente. Sua previsão é ferramenta importante em setores como as indústrias aeroespacial e de comunicação e

também na prevenção de desastres naturais. Muitas pesquisas tem sido feitas nesse campo, naturalmente dada sua importância e além disso, por se tratar de uma variável que origina séries não lineares, não estacionárias e não normais (LI; WANG, 2017).

Neste estudo os modelos individuais e combinados obtidos encontram-se descritos na Tabela 10. O modelo ARIMA não apresentou termos autorregressivos ou indicou necessidade de diferenciações, tornando-se um modelo de médias móveis de 27 termos. Já o ANN utilizou 9 nodos na camada intermediária e 50 entradas autorregressivas.

O modelo ETS indicou um erro multiplicativo e nenhuma tendência ou sazonalidade. O de SVR usou um kernel linear, 34 entradas autorregressivas e apenas quatro vetores de suporte com uma sazonalidade de 99. Dentre os modelos combinados com etapa de treinamento, o cCB modelou as marginais dos erros através da distribuição Normal e obteve uma cópula Normal para a conjunta. O cSVR usou o kernel de Base Radial com 10 vetores de suporte.

Tabela 10 – Descrição dos modelos individuais e combinados para a série SUN.

Formalismo	Modelo	Descrição
ANN	ANN(50, 9, 1)	$P_{AR} = 43, P_{ARS} = 7, S = 81, LA = BACKPROP, FA = LF$
ARIMA	ARIMA(0, 0, 27)	-
ETS	ETS(M, N, N)	$\alpha = 0.527$
SVR	-	$P_{AR} = 34, P_{ARS} = 26, S = 99, cost = 35.4373, \epsilon = 0.7972, kernel = LK, SV = 4$
cCB	-	$E_{ANN} \sim N(-6 \times 10^{-4}, sd = 9.3 \times 10^{-3}),$ $E_{ARIMA} \sim N(6, 9 \times 10^{-3}, sd = 0.123),$ $E_{ETS} \sim N(-4 \times 10^{-4}, sd = 9.8 \times 10^{-3}),$ $E_{SVR} \sim N(-1.4 \times 10^{-3}, sd = 4.8 \times 10^{-3}),$ <i>Cópula</i> = <i>Normal</i> ( $\rho_1 = 0.641$ ) $\rho_2 = 0.8474,$ $\rho_3 = 0.8759, \rho_4 = 0.829, \rho_5 = 0.4196, \rho_6 = 0.7928$ )
cSVR	-	$cost = 12.148, \epsilon = 2 \times 10^{-4}, kernel = RBFK, SV = 10$ $\gamma = 7857.7215, coef0 = 9.6558$

Fonte: Autor (2020).

Observa-se que o melhor modelo simples foi o ANN (Tabela 11), sendo SVR o pior modelo individual, considerado assim, mesmo diante da média da série, uma vez que  $ARV > 1$ . Entre as combinações, a cSM foi melhor e a pior foi a cSVR, sendo pior que um modelo de *random walk*, com *Theil* > 1.

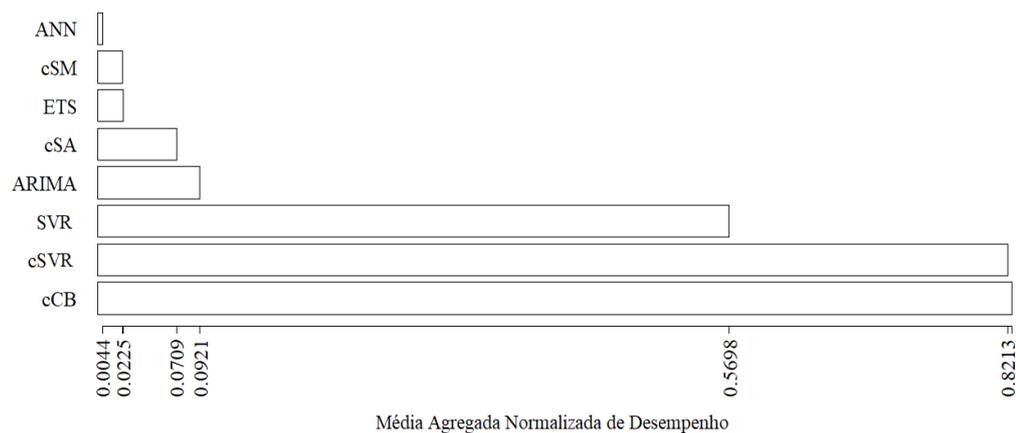
Tabela 11 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série SUN (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	<b>532.4</b>	987.3	553.3	6817.7	1033.4	622.9	15211.6	7844.3
MAPE	0.402	0.730	0.445	1.249	0.433	<b>0.401</b>	1.778	1.649
ARV	<b>0.127</b>	0.198	0.130	2.007	0.342	0.157	1.809	2.029
ID	<b>0.031</b>	0.054	0.032	0.462	0.072	0.037	0.646	0.503
Theil	<b>0.825</b>	1.530	0.858	10.571	1.602	0.966	23.586	12.162
WPOCID	<b>0.541</b>	0.581	0.589	0.546	0.583	0.577	0.631	0.998
Reg_Intercept	4.065	20.751	2.883	22.936	<b>0.203</b>	5.426	98.395	88.643
Reg_Slope	<b>0.995</b>	0.906	0.967	1.837	1.244	1.025	-1.912	0000
Reg_WR2	<b>0.112</b>	0.163	0.118	0.828	0.122	0.121	0.418	1.000

Fonte: Autor (2020).

A Figura 23 a seguir mostra a média normalizada das métricas de desempenho para a série SUN, corroborando com o resultado anterior, podemos verificar que o melhor modelo geral foi o ANN e o pior, cCB.

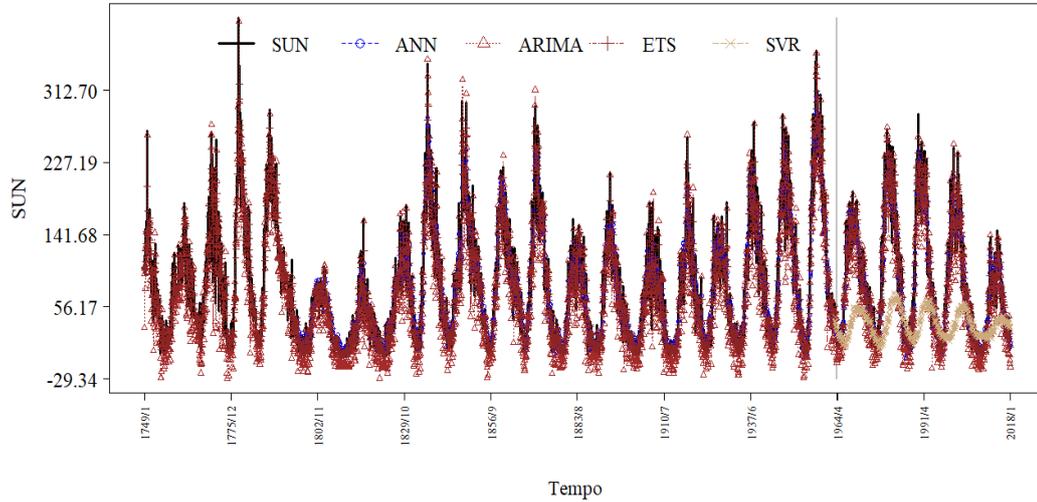
Figura 23 – Média Agregada de Performance dos modelos individuais e combinados para a série SUN.



Fonte: Autor (2020).

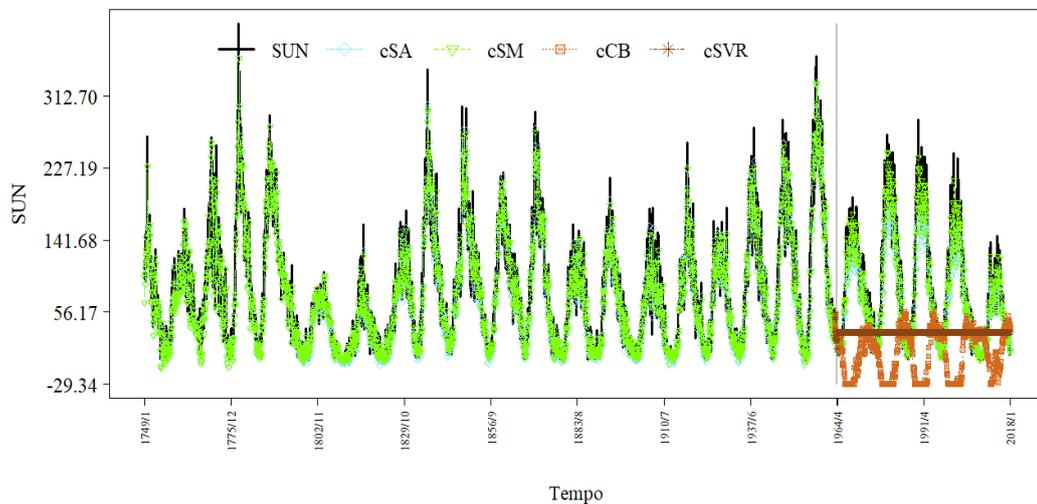
Uma apresentação visual da série SUN e seus modelos preditivos individuais e combinados podem ser observados nas Figuras 24 e 25, respectivamente.

Figura 24 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série SUN (Fase de teste).



Fonte: Autor (2020).

Figura 25 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série SUN (Fase de teste).



Fonte: Autor (2020).

Estudos como o de [Domingos, Oliveira e Neto \(2019\)](#) mostram bons resultados na previsão da série SUN utilizando um sistema híbrido com modelo ARIMA, modelos MLP e SVR com baixo número de entradas.

## 4.2 Séries Temporais Econômicas

### 4.2.1 Produção Brasileira de Biodiesel

Os biocombustíveis são derivados de biomassa renovável e podem ser uma alternativa ecologicamente mais adequada na substituição parcial ou total dos combustíveis fósseis. Os dois tipos mais comuns no Brasil são o álcool, derivado da cana-de-açúcar, e o biodiesel, produzido a partir de biomassas de origem vegetal ou animal. Atualmente, cerca de 18% dos combustíveis produzidos no Brasil já são renováveis (ANP, 2019).

A modelagem e previsão para a produção de biodiesel brasileira pode ser encontrada no trabalho de Xavier et al. (2018) que utilizou a abordagem Box e Jenkins e o alisamento exponencial de Holt-Winters, ambas as metodologias apresentaram bons resultados.

Na Tabela 12 são mostrados os modelos individuais e combinados para a série BBP. O modelo ARIMA apresentou ordem de integração igual a um, necessária para estabilizar a série, e um termo de média móvel. Já o ANN usou uma entrada autorregressiva e um nó na camada intermediária para realizar a previsão de um passo a frente. O modelo ETS diagnosticou apenas o erro multiplicativo e o SVR usou o kernel liner com 6 vetores de suporte, este preditor foi o melhor entre os individuais e também dentre todos.

Tabela 12 – Descrição dos modelos individuais e combinados para a série BBP.

Formalismo	Modelo	Descrição
ANN	ANN(1, 1, 1)	$P_{AR} = 1, P_{ARS} = 0, LA = SLR, FA = THF$
ARIMA	ARIMA(0, 1, 1)	-
ETS	ETS(M, N, N)	$\alpha = 0,735$
SVR	-	$P_{AR} = 1, P_{ARS} = 0, cost = 70.2829, \epsilon = 0.371,$ $kernel = LK, SV = 6$
cCB	-	$E_{ANN} \sim N(-5 \times 10^{-4}, sd = 9.3 \times 10^{-3}),$ $E_{ARIMA} \sim N(1.8 \times 10^{-3}, sd = 8.7 \times 10^{-3}),$ $E_{ETS} \sim N(1.8 \times 10^{-3}, sd = 8.7 \times 10^{-3}),$ $E_{SVR} \sim NormS(média = -2.4 \times 10^{-3}, sd = 0.011,$ $\xi = 1.4911), Cópula = Clayton(\alpha = 2.8673)$
cSVR	-	$cost = 98,0329, \epsilon = 0,4882, kernel = LK, SV = 3$

Fonte: Autor (2020).

Dentre os preditores combinados, cSM foi o melhor, enquanto a cCB foi o pior. Na Tabela 13 podem ser observadas as métricas de desempenho dos modelos individuais e combinados para a série BBP.

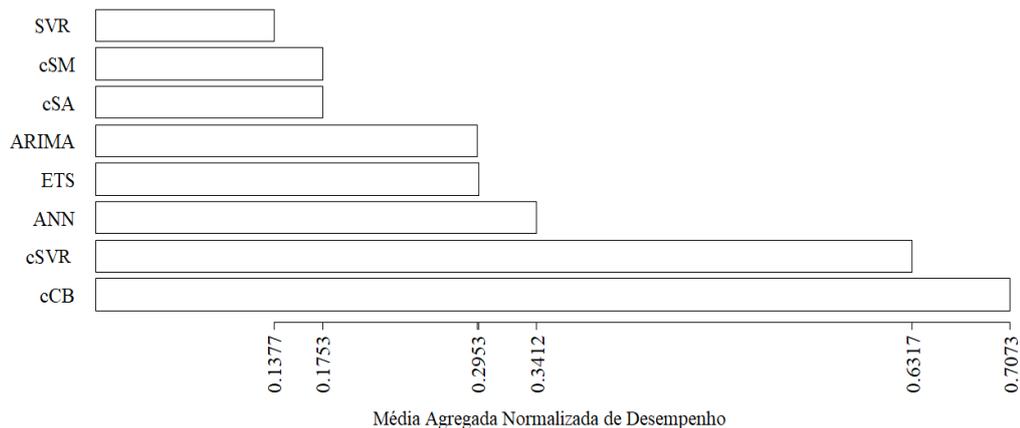
Tabela 13 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série BBP (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	831770	765839	766039	744696	745838	<b>742232</b>	889378	971081
MAPE	0.070	0.071	0.071	<b>0.068</b>	0.068	0.069	0.077	0.073
ARV	1.041	<b>0.708</b>	0.709	0.714	0.769	0.715	0.732	1.061
ID	0.201	0.160	0.161	0.158	0.164	<b>0.158</b>	0.177	0.227
Theil	1.128	1.045	1.046	1.014	1.016	<b>1.012</b>	1.216	1.315
WPOCID	<b>0.536</b>	<b>0.536</b>	<b>0.536</b>	<b>0.536</b>	<b>0.536</b>	<b>0.536</b>	<b>0.536</b>	<b>0.536</b>
Reg_Intercept	<b>25625</b>	59568	59522	54488	48182	54971	78419	37940
Reg_Slope	<b>0.959</b>	0.829	0.829	0.858	0.873	0.850	0.764	0.935
Reg_WR2	0.419	0.449	0.450	<b>0.410</b>	0.428	0.427	0.504	0.417

Fonte: Autor (2020).

A Figura 26 mostra a média das métricas de desempenho normalizadas, na qual pode-se verificar que o melhor modelo geral foi o SVR e o pior foi o cCB. Nenhum modelo apresentaram desempenho melhor que um modelo de *random walk* (*Theil* > 1). É possível notar que todos os modelos apresentaram MSE muito elevados, também devido à magnitude dos valores da série sob estudo.

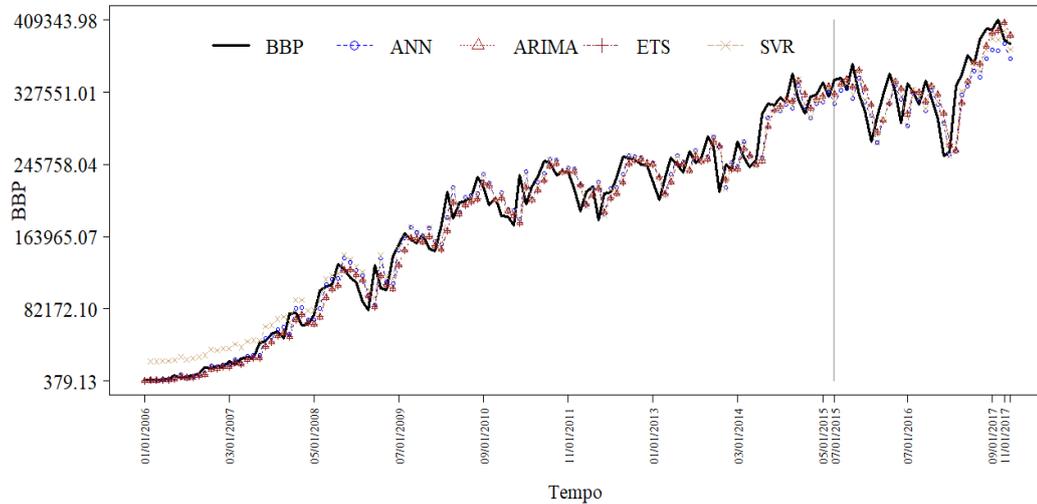
Figura 26 – Média Agregada de Performance dos modelos individuais e combinados para a série BBP.



Fonte: Autor (2020).

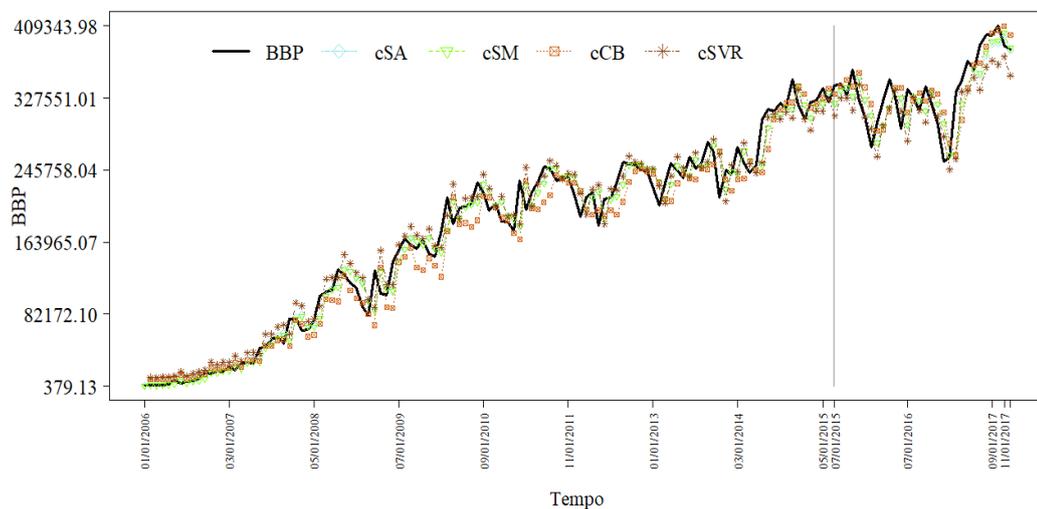
Nas Figuras 27 e 28 estão as representações gráficas das previsões para a série BBP.

Figura 27 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série BBP (Fase de teste).



Fonte: Autor (2020).

Figura 28 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série BBP (Fase de teste).



Fonte: Autor (2020).

#### 4.2.2 Índice de Preços ao Consumidor no Brasil

O IPCA é aferido mensalmente pelo Instituto Brasileiro de Geografia e Estatística – IBGE com a finalidade de oferecer ao consumidor final a variação dos preços do comércio. Este índice é utilizado pelo governo brasileiro como índice oficial de inflação, importante

variável de governança relativa ao aumento dos preços de produtos e serviços (IBGE, 2019).

A Tabela 14 mostra a arquitetura dos modelos individuais e combinados para a série IPCA. O modelo ARIMA apresentou apenas termos de médias móveis, o ETS é uma suavização exponencial simples de erro multiplicativo.

Tabela 14 – Descrição dos modelos individuais e combinados para a série IPCA.

Formalismo	Modelo	Descrição
ANN	ANN(1, 1, 1)	$P_{AR} = 1, P_{ARS} = 0, LA = RPROP^-, FA = LF$
ARIMA	ARIMA(0, 0, 6)	-
ETS	ETS(M, N, N)	$\alpha = 0.878$
SVR	-	$P_{AR} = 3, P_{ARS} = 3, S = 11, cost = 78.6895, \epsilon = 0.9542, kernel = LK, SV = 13$
cCB	-	$E_{ANN} \sim N(-1.9 \times 10^{-3}, sd = 0.0119),$ $E_{ARIMA} \sim N(0.0342, sd = 0.0301),$ $E_{ETS} \sim N(-3 \times 10^{-4}, sd = 0.0144),$ $E_{SVR} \sim N(5 \times 10^{-4}, sd = 0.0126),$ <i>Cópula = Normal</i> ( $\rho_1 = 0.3479, \rho_2 = 0.9005,$ $\rho_3 = 0.9608, \rho_4 = 0.3683, \rho_5 = 0.373, \rho_6 = 0.9438$ )
cSVR	-	$cost = 95.5887, \epsilon = 0.8413, kernel = PK, SV = 21$ $\gamma = 4254.5389, degree = 2.3027, coef0 = 5.5671$

Fonte: Autor (2020).

Na Tabela 15 encontra-se um resumo das métricas de desempenho dos modelos individuais e combinados para a série IPCA. O modelo ANN apresentou melhor MSE e Theil, sendo melhor em detrimento do modelo de *random walk*. O modelo individual de melhor desempenho foi o ANN e o de pior, o ARIMA, entretanto foi o que mais acertou a previsão da tendência da série. O melhor preditor combinado foi cCB, que modelou as distribuições marginais dos erros através de Normais e a conjunta com uma cópula Normal.

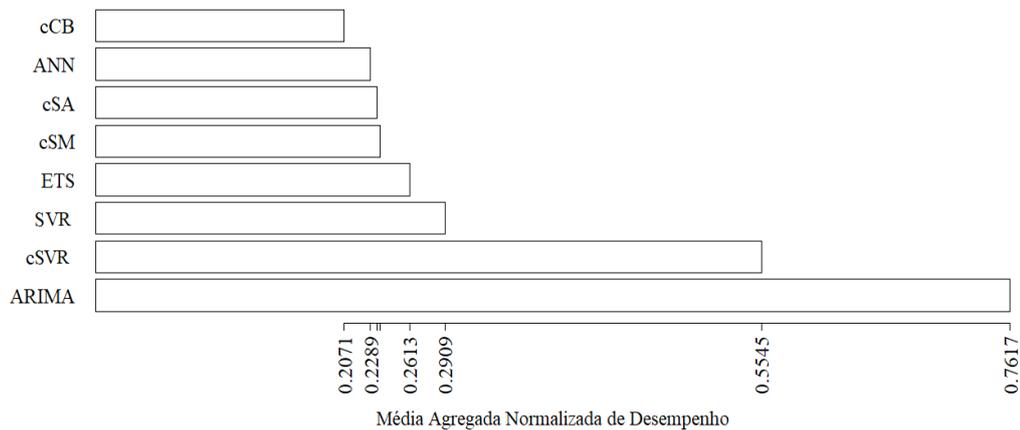
Tabela 15 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série IPCA (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	<b>0.080</b>	0.646	0.097	0.092	0.120	0.090	0.082	0.182
MAPE	0.796	2.194	0.791	0.810	<b>0.746</b>	0.802	0.766	1.258
ARV	1.344	0.691	0.984	1.006	<b>0.640</b>	0.974	1.123	1.344
ID	0.272	0.421	0.263	0.266	<b>0.230</b>	0.256	0.254	0.474
Theil	<b>0.761</b>	6.267	0.942	0.889	1.157	0.871	0.772	1.744
WPOCID	0.682	<b>0.568</b>	0.682	0.705	0.659	0.659	0.682	0.614
Reg_Intercept	<b>-0.013</b>	0.486	0.172	0.141	0.268	0.140	-0.033	0.245
Reg_Slope	<b>0.965</b>	0.218	0.620	0.705	0.578	0.719	1.073	0.396
Reg_WR2	0.620	0.819	<b>0.616</b>	0.665	0.665	0.658	0.643	0.756

Fonte: Autor (2020).

A Figura 29 ilustra a média das métricas de desempenho normalizadas, na qual pode-se verificar que o desempenho do modelo ARIMA foi o pior geral e o cCB foi o melhor.

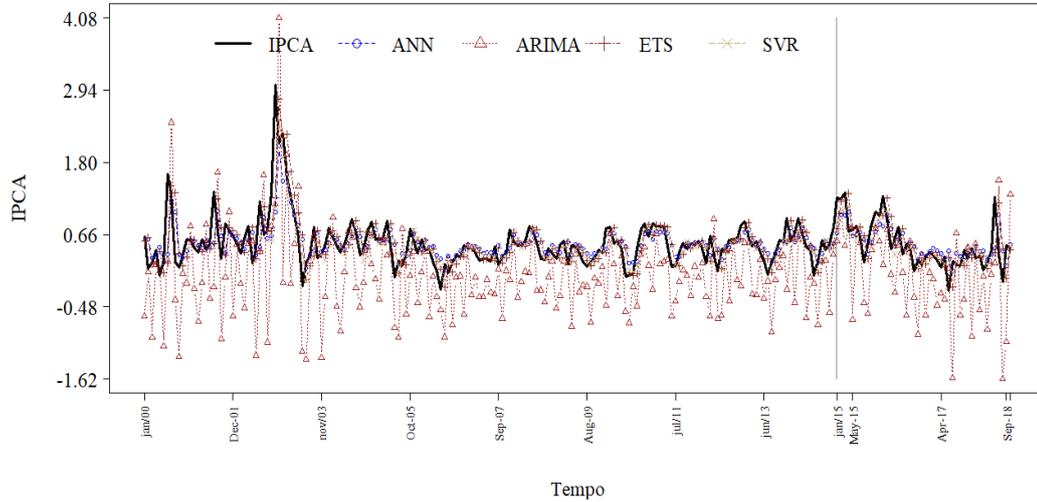
Figura 29 – Média Agregada de Performance dos modelos individuais e combinados para a série IPCA.



Fonte: Autor (2020).

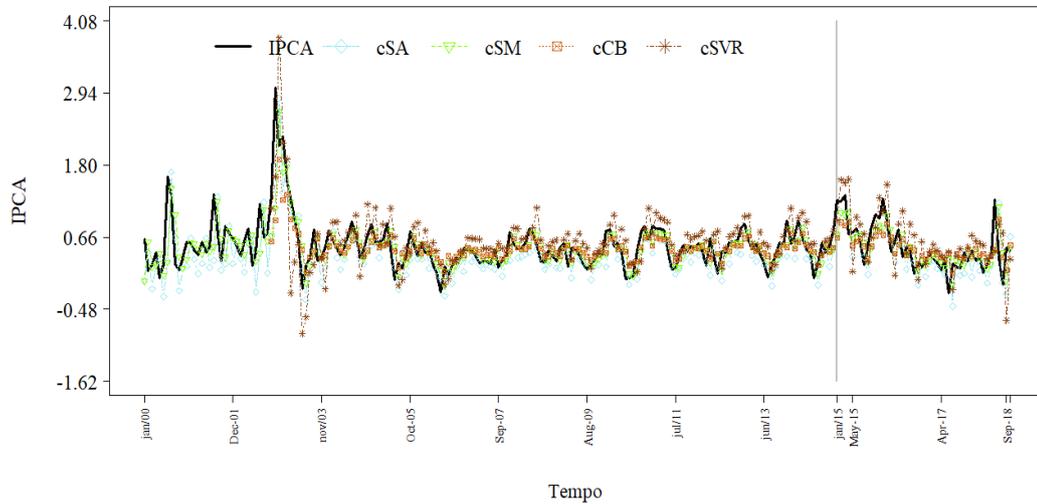
Nas Figuras 30 e 31 a seguir, são mostrados graficamente os modelos individuais e combinados para a série IPCA.

Figura 30 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série IPCA (Fase de teste).



Fonte: Autor (2020).

Figura 31 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série IPCA (Fase de teste).



Fonte: Autor (2020).

#### 4.2.3 Preço de ações da Petrobras

A previsão da direção das mudanças nos preços das ações contribui para o desenvolvimento de estratégias adequadas em transações nos mercados financeiros. Nesse sentido, tem sido comum a busca por modelos matemáticos ou algoritmos computacionais capazes de estimar as flutuações nos preços (OLIVEIRA; NOBRE; ZARATE, 2013).

A Petrobras é uma sociedade de capital aberto que atua na indústria de óleo, gás natural e energia. Atualmente, as ações da Petrobras representam uma das maiores empresas do mercado de ações brasileiro (PETROBRAS, 2019; NETO; CALVALCANTI; REN, 2009).

A Tabela 16 mostra a estrutura dos modelos individuais e combinados para a série PSP. O modelo ARIMA exigiu apenas uma diferenciação, sendo equivalente a um *random walk* ( $Theil = 1$ ), assim como o ETS. No modelo ANN foi considerada uma entrada autorregressiva e um nó na camada intermediária.

Tabela 16 – Descrição dos modelos individuais e combinados para a série PSP.

Formalismo	Modelo	Descrição
ANN	ANN(1, 1, 1)	$P_{AR} = 1, P_{ARS} = 0, LA = SLR, FA = THF$
ARIMA	ARIMA(0, 1, 0)	-
ETS	ETS(M, N, N)	$\alpha = 1.0$
SVR	-	$P_{AR} = 3, P_{ARS} = 0, cost = 20.1788, \epsilon = 0.9841, kernel = LK, SV = 3$
cCB	-	$E_{ANN} \sim N(-4 \times 10^{-4}, sd = 7 \times 10^{-3}),$ $E_{ARIMA} \sim N(-2 \times 10^{-4}, sd = 7.1 \times 10^{-3}),$ $E_{ETS} \sim N(-2 \times 10^{-4}, sd = 7.1 \times 10^{-3}),$ $E_{SVR} \sim N(6.7 \times 10^{-3}, sd = 7.2 \times 10^{-3}),$ $Cópula = Gumbel(\alpha = 6.5394)$
cSVR	-	$cost = 2.0518, \epsilon = 0.9449, kernel = LK, SV = 2$

Fonte: Autor (2020).

O ETS foi o melhor modelo individual. O cCB apresentou menor MSE, sendo o melhor preditor combinado, acertou mais a previsão da tendência da série (*WPOCID*) e foi o melhor entre todos. Já o SVR simples teve o pior resultado individual, como mostra a Tabela 17.

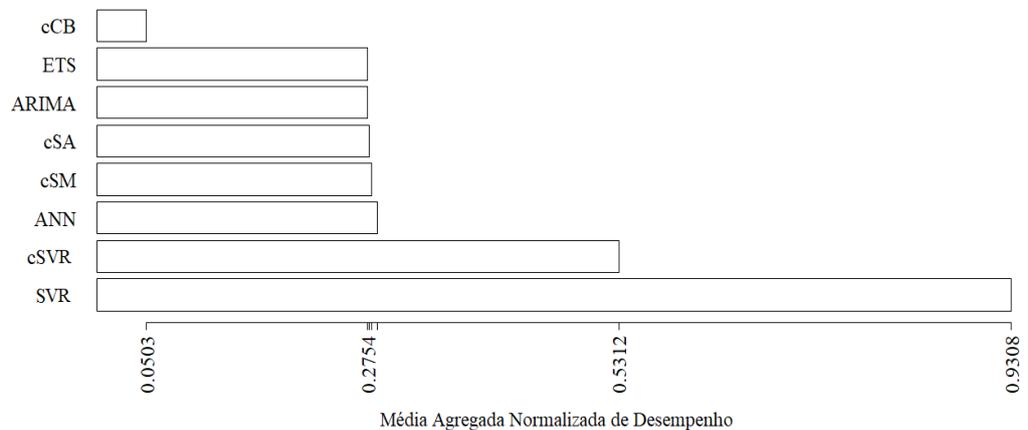
Tabela 17 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série PSP (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	4.104	4.209	4.209	7.383	4.487	4.214	<b>3.905</b>	6.139
MAPE	0.119	<b>0.117</b>	0.117	0.149	0.118	0.117	0.118	0.133
ARV	0.151	0.168	0.168	0.474	0.197	0.168	<b>0.144</b>	0.347
ID	0.037	0.040	0.040	0.088	0.044	0.040	<b>0.035</b>	0.069
Theil	0.974	1.000	1.000	1.746	1.066	1.001	<b>0.922</b>	1.454
WPOCID	0.452	0.452	0.452	0.476	0.429	0.452	<b>0.405</b>	0.429
Reg_Intercept	0.550	0.630	0.630	1.575	0.824	0.639	<b>0.197</b>	1.409
Reg_Slope	0.966	0.977	0.977	1.014	0.984	0.976	0.986	<b>1.001</b>
Reg_WR2	0.131	0.133	0.133	0.140	0.133	0.133	<b>0.126</b>	0.133

Fonte: Autor (2020).

O pior modelo para a série PSP foi o SVR simples, como ilustra a Figura 32, que usou um kernel linear e apenas 2 vetores de suporte.

Figura 32 – Média Agregada de Performance dos modelos individuais e combinados para a série PSP.

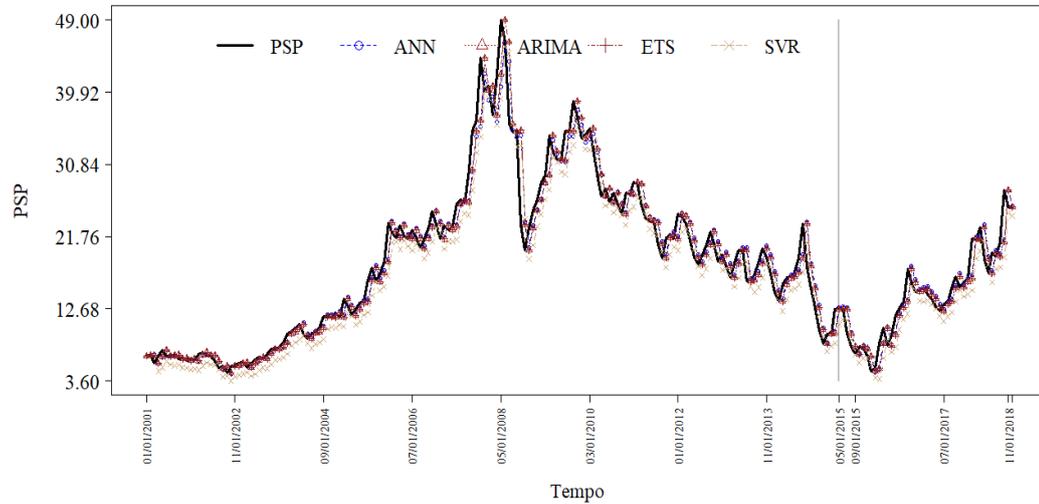


,m4589

Fonte: Autor (2020).

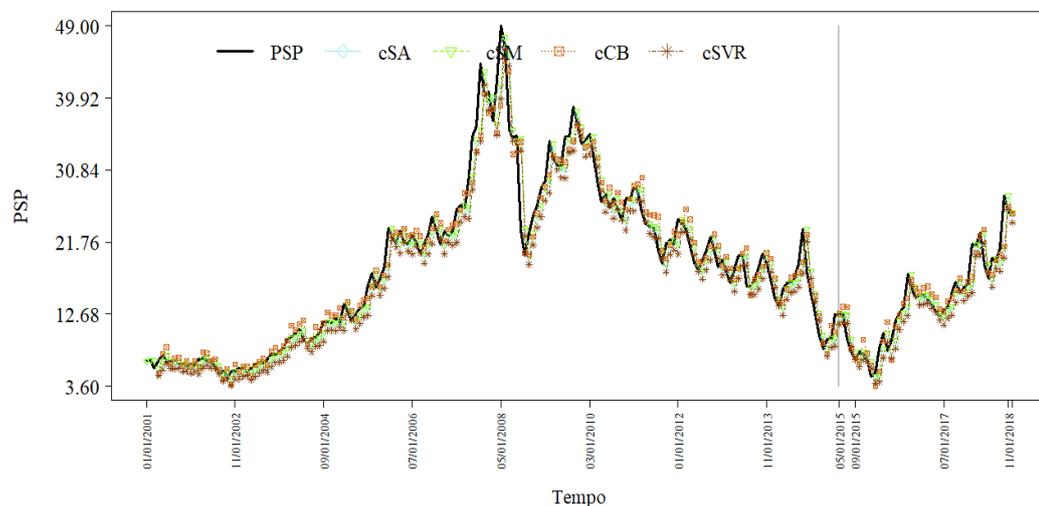
Nas Figuras 33 e 34 pode-se verificar as previsões dos modelos individuais e combinados para a série PSP graficamente.

Figura 33 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série PSP (Fase de teste).



Fonte: Autor (2020).

Figura 34 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série PSP (Fase de teste).



Fonte: Autor (2020).

#### 4.2.4 Índice *Standard & Poor's 500*

O índice S&P 500 é um indicador do mercado de ações ponderado pela capitalização de mercado, ajustado por flutuação livre nos Estados Unidos. Seu valor é influenciado

por outros importantes índices financeiros, como o preço de commodities e indicadores técnicos financeiros (LIU et al., 2016).

Muitos trabalhos têm utilizado a abordagem de séries temporais para previsão do movimento no mercado de ações, entre os quais Rout et al. (2017) utilizou uma metodologia baseada em redes neurais recorrentes e algoritmos de aprendizado evolutivo para a previsão do S&P 500.

A Tabela 18 mostra os modelos individuais e combinados para a série SP500. O modelo ARIMA apresentou apenas uma diferenciação e o ETS, uma suavização exponencial simples de erro multiplicativo. A cSVR utilizou o kernel linear e 6 vetores de suporte. Na cCB as distribuições marginais e conjunta dos erros foram Normais.

Tabela 18 – Descrição dos modelos individuais e combinados para a série SP500.

Formalismo	Modelo	Descrição
ANN	ANN(28, 5, 1)	$P_{AR} = 20, P_{ARS} = 8, S = 17, LA = SLR, FA = LHF$
ARIMA	ARIMA(0, 1, 0)	-
ETS	ETS(M, N, N)	$\alpha = 0.942$
SVR	-	$P_{AR} = 3, P_{ARS} = 0, cost = 17.885, \epsilon = 0.275,$ $kernel = LK, SV = 6$
cCB	-	$E_{ANN} \sim N(4 \times 10^{-4}, sd = 4.7 \times 10^{-3}),$ $E_{ARIMA} \sim N(5.4 \times 10^{-4}, sd = 2.2 \times 10^{-3}),$ $E_{ETS} \sim N(5 \times 10^{-4}, sd = 2.2 \times 10^{-3}),$ $E_{SVR} \sim N(1.3 \times 10^{-3}, sd = 3.1 \times 10^{-3}),$ $Cópula = Normal (\rho_1 = 0.5779) \rho_2 = 0.6065,$ $\rho_3 = 0.644, \rho_4 = 0.9954, \rho_5 = 0.7949, \rho_6 = 0.8094)$
cSVR	-	$cost = 5.1625, \epsilon = 0.6071, kernel = LK, SV = 2$

Fonte: Autor (2020).

A Tabela 19 mostra o desempenho dos preditores para a série SP500. O ETS foi o único modelo melhor que o modelo de *random walk* ( $Theil < 1$ ), sendo o melhor entre todos os preditores. Apenas os modelos ANN e cSVR não foram melhores quando comparados com a média dos valores regressos da série ( $ARV > 1$ ).

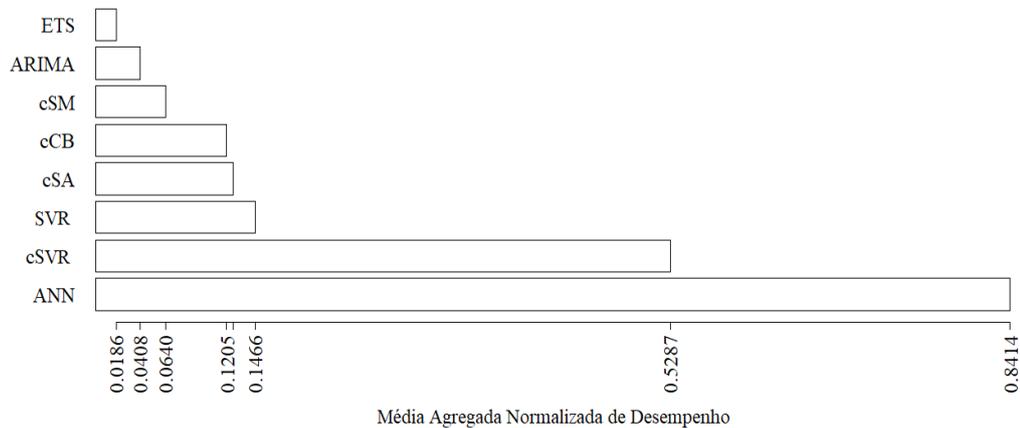
Tabela 19 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série SP500 (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	12094.5	503.2	<b>501.1</b>	1072.5	1587.6	650.9	559.6	5664.7
MAPE	0.033	0.006	<b>0.006</b>	0.010	0.012	0.007	0.006	0.024
ARV	1.730	0.042	0.042	0.126	0.202	0.065	<b>0.042</b>	1.189
ID	0.370	0.010	<b>0.010</b>	0.026	0.041	0.015	0.011	0.190
Theil	24.040	1.000	<b>0.996</b>	2.132	3.156	1.294	1.112	11.261
WPOCID	0.500	0.480	0.475	0.490	<b>0.470</b>	0.480	0.495	<b>0.470</b>
Reg_Intercept	173.70	31.7	<b>25.7</b>	-72.9	-104.3	-46.3	64.0	-366.4
Reg_Slope	0.972	0.990	<b>0.992</b>	1.033	1.046	1.020	0.976	1.150
Reg_WR2	0.258	0.040	<b>0.040</b>	0.040	0.051	0.040	0.041	0.049

Fonte: Autor (2020).

Na Figura 35 pode-se observar o desempenho dos preditores, o de pior resultado geral e individual foi o ANN e a pior combinação foi cSVR. O melhor modelo combinado foi a cSM.

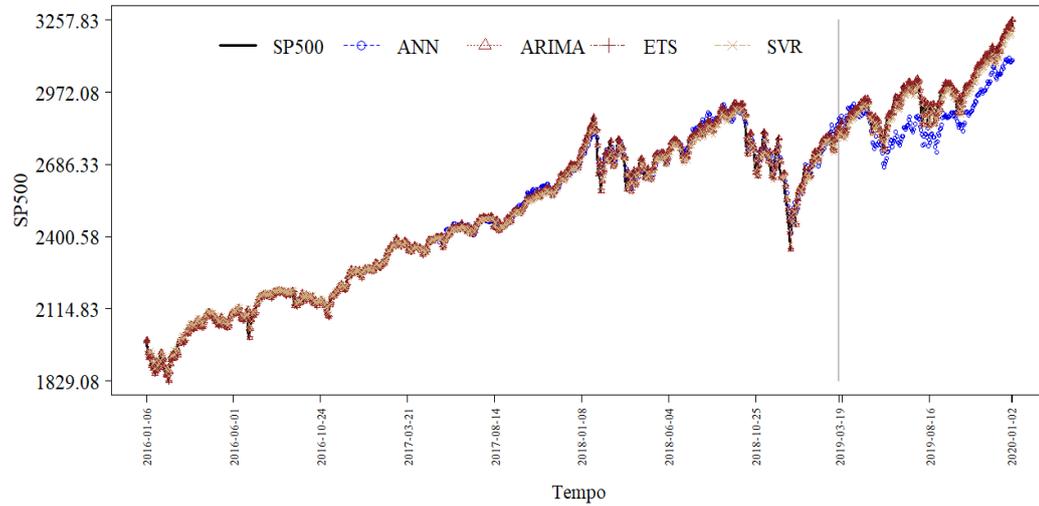
Figura 35 – Média Agregada de Performance dos modelos individuais e combinados para a série SP500.



Fonte: Autor (2020).

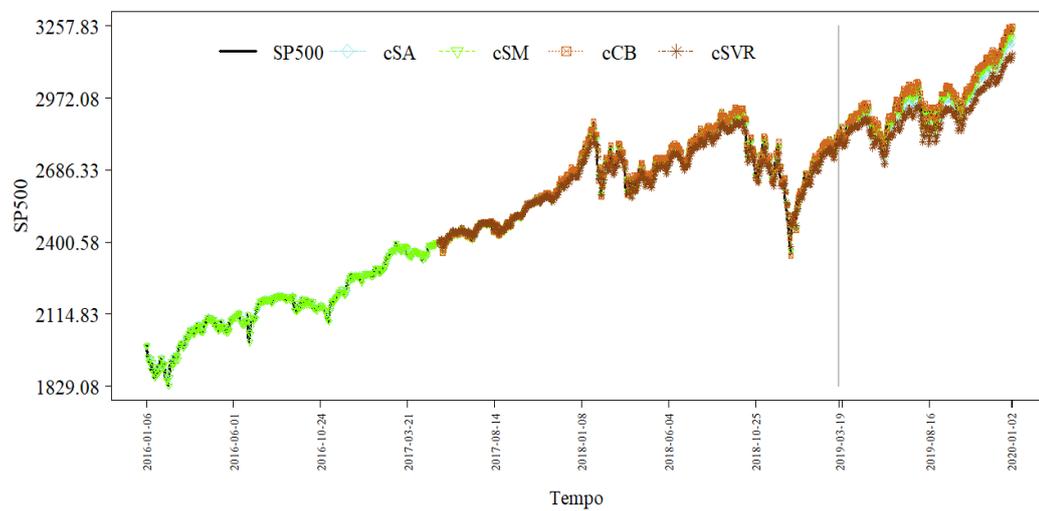
As Figuras 36 e 37 adiante mostram as previsões dos preditores simples e combinados para a série SP500.

Figura 36 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série SP500 (Fase de teste).



Fonte: Autor (2020).

Figura 37 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série SP500 (Fase de teste).



Fonte: Autor (2020).

## 4.3 Séries Temporais Sociais

### 4.3.1 Mortes por Veículos Motorizados

Os acidentes de trânsito representam um grave problema de saúde pública a nível global. O Brasil é um dos países com maior número de vítimas fatais no trânsito, precedido apenas da Índia, mesmo com uma legislação muito rigorosa comparada a outras nações (ABREU; SOUZA; MATHIAS, 2018).

A série MVF trata-se de observações do total de vítimas fatais em acidentes de trânsito no estado do Ceará-Brasil. A descrição dos modelos está na Tabela 20. Tem-se um modelo  $IMA(1, 1)$  na classe ARIMA e, na ETS, uma suavização exponencial simples de erro multiplicativo. O modelo ANN usou uma entrada autorregressiva e um nó na camada intermediária.

Tabela 20 – Descrição dos modelos individuais e combinados para a série MVF.

Formalismo	Modelo	Descrição
ANN	ANN(1, 1, 1)	$P_{AR} = 1, P_{ARS} = 0, LA = SAG, FA = LHF$
ARIMA	ARIMA(0, 1, 1)	-
ETS	ETS(M, N, N)	$\alpha = 0, 345$
SVR	-	$P_{AR} = 1, P_{ARS} = 0, cost = 48.6353, \epsilon = 1,$ $kernel = LK, SV = 10,$
cCB	-	$E_{ANN} \sim N(0, sd = 0.0321),$ $E_{ARIMA} \sim N(2.4 \times 10^{-3}, sd = 0.0311),$ $E_{ETS} \sim N(2.9 \times 10^{-3}, sd = 0.031),$ $E_{SVR} \sim N(-4.6 \times 10^{-3}, sd = 0.0304),$ $Cópula = Gumbel(\alpha = 3.2097)$
cSVR	-	$cost = 33, 4079, \epsilon = 0, 9843, kernel = PK, SV = 7$ $\gamma = 6671, 2204, degree = 1, 3537 coef0 = 4, 1292$

Fonte: Autor (2020).

Segundo a Tabela 21, o modelo ARIMA foi o melhor para a série. Já o preditor SVR, modelado com kernel linear e 10 vetores de suporte, foi o pior. Entre as combinações, a cSM mostrou melhor desempenho. A pior estimativa combinada foi dada pelo cSVR, com kernel polinomial e 7 vetores de suporte. Todos os preditores mostraram  $MSE$  elevados e nenhum modelo é vantajoso comparado com a média regressa da série, uma vez que todos apresentam  $ARV > 1$ . Os resultados podem ter sido comprometidos pela pequena quantidade de observações da série.

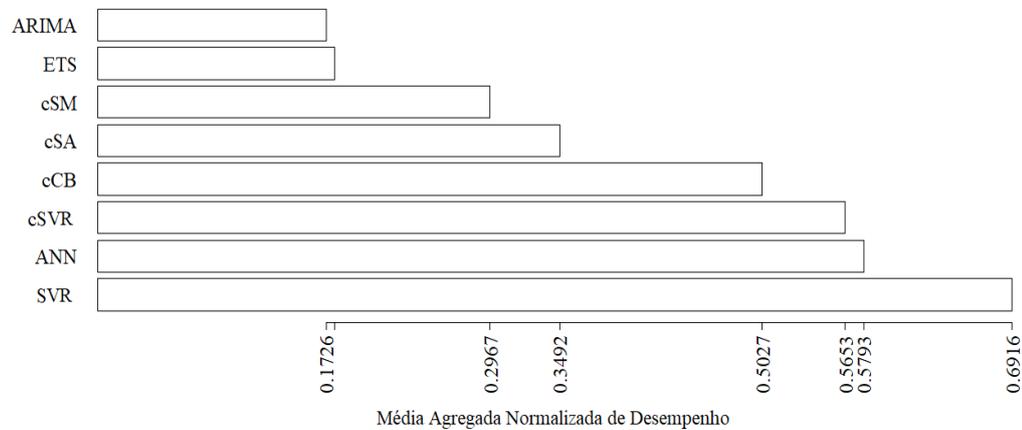
Tabela 21 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série MVF (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	1635.7	1301.9	1290.6	1733.2	<b>1421.3</b>	1396.7	1830.5	1391.1
MAPE	0.170	0.161	0.161	0.177	<b>0.160</b>	0.159	0.177	0.168
ARV	2.192	3.085	3.284	<b>2.331</b>	2.803	2.749	2.239	5.639
ID	0.582	0.595	0.608	0.604	<b>0.595</b>	0.586	0.587	0.786
Theil	0.763	0.609	0.605	0.818	<b>0.664</b>	0.652	0.899	0.664
WPOCID	0.750	0.812	0.812	0.750	0.812	0.812	<b>0.625</b>	0.625
Reg_Intercept	<b>157.2</b>	110.1	107.8	160.1	132.2	128.0	143.49	138.3
Reg_Slope	<b>0.213</b>	0.448	0.459	0.196	0.336	0.358	0.271	0.298
Reg_WR2	0.976	0.930	0.932	0.977	<b>0.955</b>	0.950	0.935	0.984

Fonte: Autor (2020).

A Figura 38 resume o desempenho dos modelos para a série MVF.

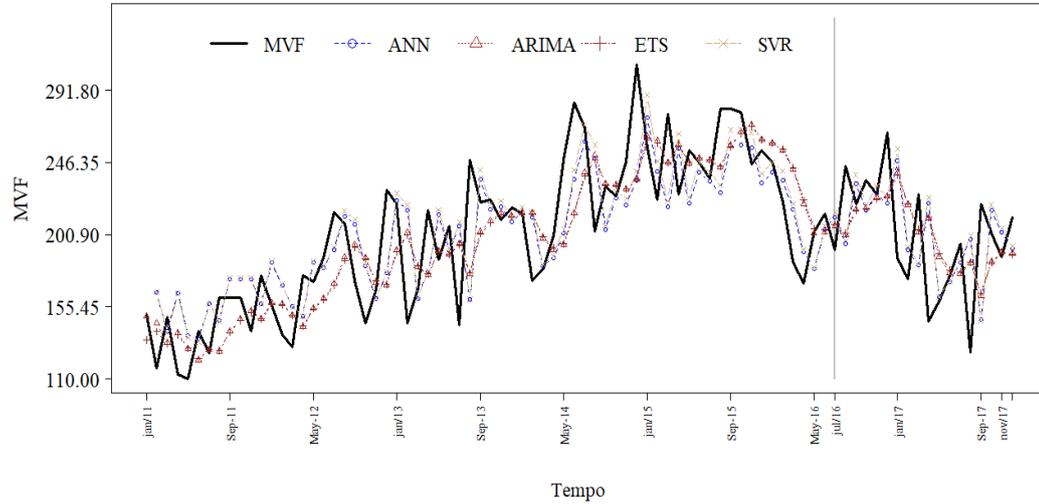
Figura 38 – Média Agregada de Performance dos modelos individuais e combinados para a série MVF.



Fonte: Autor (2020).

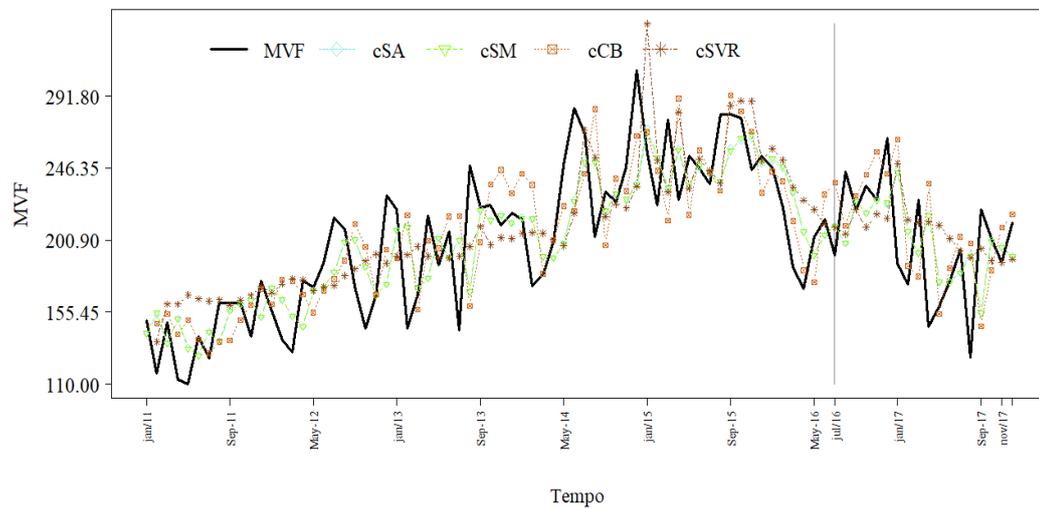
As Figuras 39 e 40 permitem uma análise gráfica do desempenho dos modelos individuais e combinados para a série MVF.

Figura 39 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série MVF (Fase de teste).



Fonte: Autor (2020).

Figura 40 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série MVF (Fase de teste).



Fonte: Autor (2020).

### 4.3.2 Número de Casos de Dengue no Rio de Janeiro

A dengue é uma doença endêmica que ocorre mais em área tropical e é de grande relevância mundial pela fácil disseminação e alta incidência. É causada pelo arbovírus do gênero flavivírus e transmitida pelo mosquito *Aedes aegypti* (CARVALHO; MAGALHÃES; MEDRONHO, 2017).

Modelos aditivos generalizados, redes neurais artificiais, modelos de médias móveis auto-regressivos sazonais (BAQUERO; SANTANA; CHIARAVALLOTI-NETO, 2018) e *machines learning* (CARVALHO; MAGALHÃES; MEDRONHO, 2017) têm sido empregados para estimar o número total de casos de dengue.

Há também pesquisas ligadas a identificar um ano epidêmico, usando SVM para a classificação (STOLERMAN; MAIA; KUTZ, 2019).

A Tabela 22 mostra os modelos para a série DEN. O ARIMA, com apenas a diferenciação de ordem 1, e o ETS, suavização exponencial simples, mostraram-se equivalentes. O modelo de SVR usou um kernel de base radial e 58 vetores de suporte. O modelo de cCB modelou as distribuições marginais dos erros dos preditores individuais através das distribuições Normal ( $N$ ), Laplace ( $L$ ) e Cauchy ( $C$ ) e a conjunta por uma cópula Frank.

Tabela 22 – Descrição dos modelos individuais e combinados para a série DEN.

Formalismo	Modelo	Descrição
ANN	ANN(3, 3, 1)	$P_{AR} = 3, P_{ARS} = 0, LA = BACKPROP, FA = THF$
ARIMA	ARIMA(0, 1, 0)	-
ETS	ETS(M, N, N)	$\alpha = 1.0$
SVR	-	$P_{AR} = 1, P_{ARS} = 0, cost = 98.1534, \epsilon = 0.0253,$ $kernel = RBFK, SV = 58, \gamma = 5498.7475$
cCB	-	$E_{ANN} \sim L(-7 \times 10^{-4}, s = 8 \times 10^{-4}),$ $E_{ARIMA} \sim C(loca\c{c}{a}{\c{c}{o}} = 1 \times 10^{-4}, escala = 5 \times 10^{-4}),$ $E_{ETS} \sim C(loca\c{c}{a}{\c{c}{o}} = 1 \times 10^{-4}, escala = 5 \times 10^{-4}),$ $E_{SVR} \sim N(-1 \times 10^{-4}, sd = sd = 1 \times 10^{-3}),$ $C\acute{o}pula = Frank(\alpha = 5.3451)$
cSVR	-	$cost = 23.5802, \epsilon = 0.055, kernel = LK, SV = 6$

Fonte: Autor (2020).

ARIMA e ETS foram os modelos mais eficientes. Apesar de bons, esses modelos são equivalentes aos modelo de *randon walk* ou o que retorna a média da série, com *Theil* e *ARV* maiores que um. A cSVR foi o pior entre os modelos combinados. SVR teve o pior desempenho dentre os modelos individuais (Tabela 23).

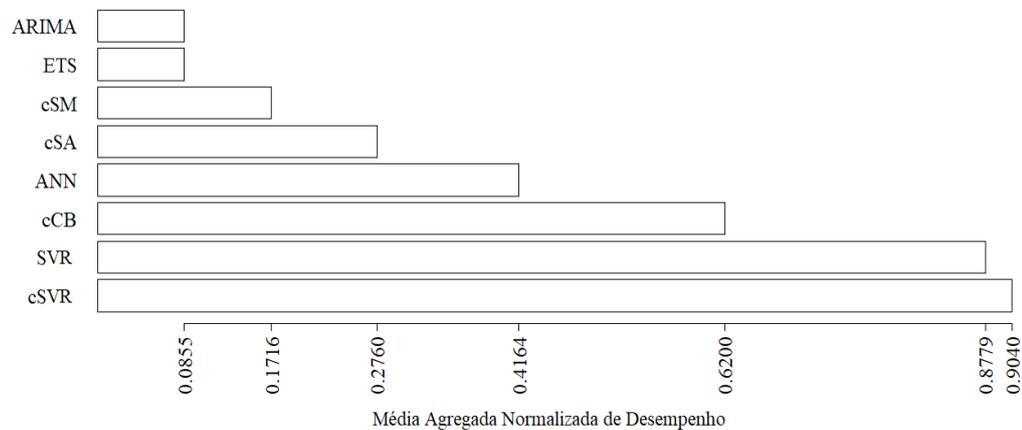
Tabela 23 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série DEN (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	31254	14412	<b>14412</b>	132899	34769	21580	91556	129030
MAPE	0.947	<b>0.517</b>	0.517	0.977	0.612	0.551	0.751	1.062
ARV	<b>0.880</b>	1.002	1.002	1.045	1.011	1.003	1.021	1.047
ID	0.680	<b>0.647</b>	0.647	0.951	0.818	0.735	0.925	0.954
Theil	1.329	1.000	1.000	10.966	1.204	<b>0.987</b>	2.872	10.232
WPOCID	0.458	<b>0.375</b>	<b>0.375</b>	0.458	<b>0.375</b>	<b>0.375</b>	0.417	0.500
Reg_Intercept	274.93	229.63	<b>229.63</b>	302.73	255.66	244.29	313.09	293.11
Reg_Slope	0.252	<b>0.347</b>	0.347	0.144	0.254	0.302	0.168	0.146
Reg_WR2	0.202	0.095	0.095	0.151	0.092	<b>0.091</b>	0.130	0.146

Fonte: Autor (2020).

Na Figura 41 pode-se observar que o melhor combinador para o caso foi o cSM.

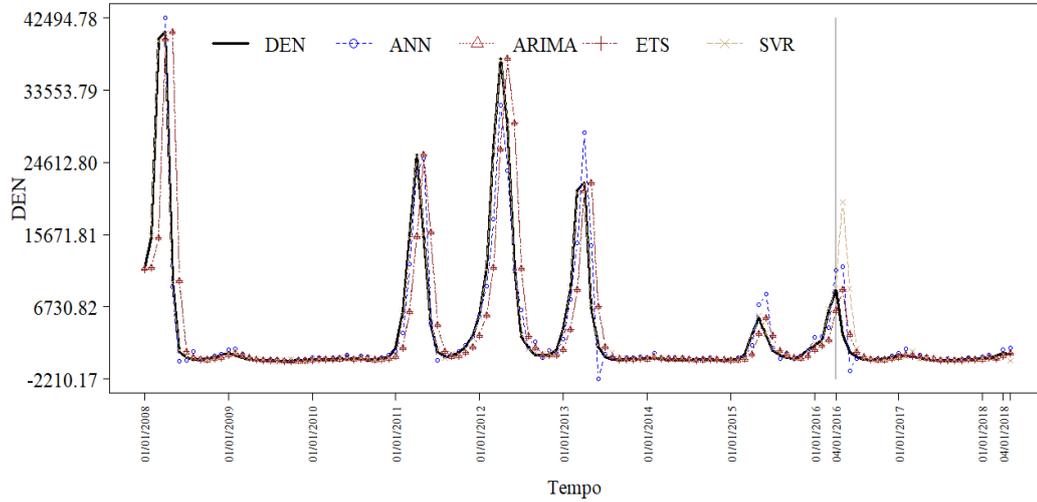
Figura 41 – Média Agregada de Performance dos modelos individuais e combinados para a série DEN.



Fonte: Autor (2020).

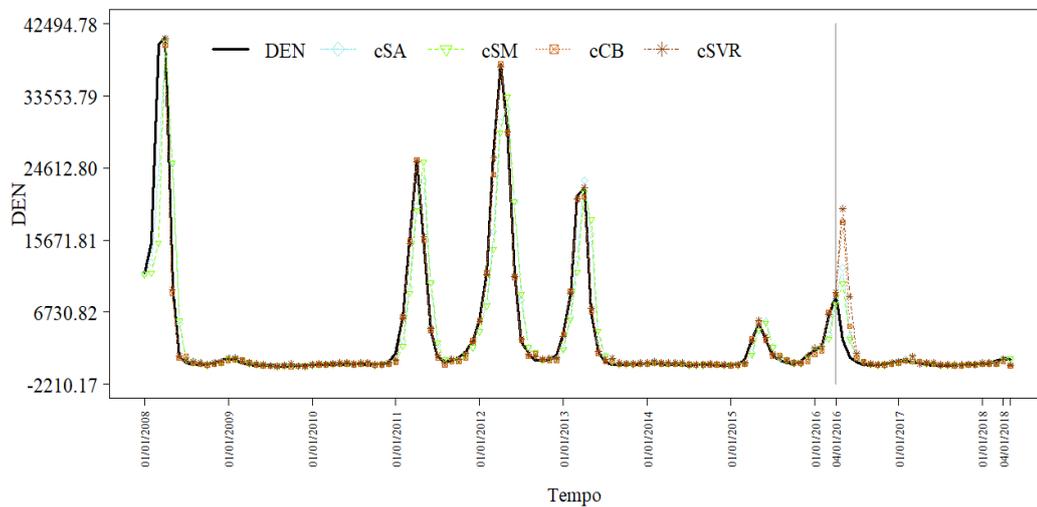
As Figuras 42 e 43 permitem uma análise visual do desempenho dos preditores individuais e combinados para a série DEN.

Figura 42 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série DEN (Fase de teste).



Fonte: Autor (2020).

Figura 43 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série DEN (Fase de teste).



Fonte: Autor (2020).

### 4.3.3 Consumo de Energia no Nordeste Brasileiro

Nas economias em desenvolvimento é natural um aumento no consumo de energia, dessa forma, para melhorar o suprimento de energia é de suma importância estimar a sua demanda (BARAK; SADEGH, 2016).

Muitos estudiosos dedicaram-se a prever a demanda de energia elétrica. Por exemplo, [Zhang et al. \(2016\)](#) utilizou Regressão de vetores de suporte para estimar a demanda de energia de um edifício. Modelos da classe ARIMA foram empregados por [Ozturk e Ozturk \(2018\)](#).

A Tabela 24 mostra os modelos para a série ENE. Tem-se um modelo de médias móveis com uma diferenciação e uma suavização exponencial simples com erro multiplicativo. Os modelos ANN e SVR usaram 3 e 2 entradas autorregressivas, nesta ordem. Os modelos SVR e cSVR utilizaram apenas 2 vetores de suporte.

Tabela 24 – Descrição dos modelos individuais e combinados para a série ENE.

Formalismo	Modelo	Descrição
ANN	ANN(3, 2, 1)	$P_{AR} = 3, P_{ARS} = 0, S = 3, LA = SLR, FA = THF$
ARIMA	ARIMA(0, 1, 1)	-
ETS	ETS(M, N, N)	$\alpha = 0.762$
SVR	-	$P_{AR} = 2, P_{ARS} = 0, S = 15, cost = 64.8515, \epsilon = 0.5673,$ $kernel = PK, SV = 2$ $\gamma = 6597.6, degree = 1.5458, coef0 = 5.0555$
cCB	-	$E_{ANN} \sim N(5 \times 10^{-4}, sd = 5.1 \times 10^{-3}),$ $E_{ARIMA} \sim N(1.3 \times 10^{-3}, sd = 4.5 \times 10^{-3}),$ $E_{ETS} \sim N(1.3 \times 10^{-3}, sd = 4.5 \times 10^{-3}),$ $E_{SVR} \sim N(9.7 \times 10^{-3}, sd = 4.7 \times 10^{-3}),$ $Cópula = Frank(\alpha = 13.6674)$
cSVR	-	$cost = 84.2117, \epsilon = 0.522, kernel = LK, SV = 2$

Fonte: Autor (2020).

O modelo ETS foi o de melhor desempenho entre todos os modelos. O modelo cCB teve distribuições marginais dos erros Normais e conjunta, Frank. Na Tabela 25 traz dados dos desempenhos dos modelos para a série ENE. Nota-se que o SVR teve o pior desempenho geral.

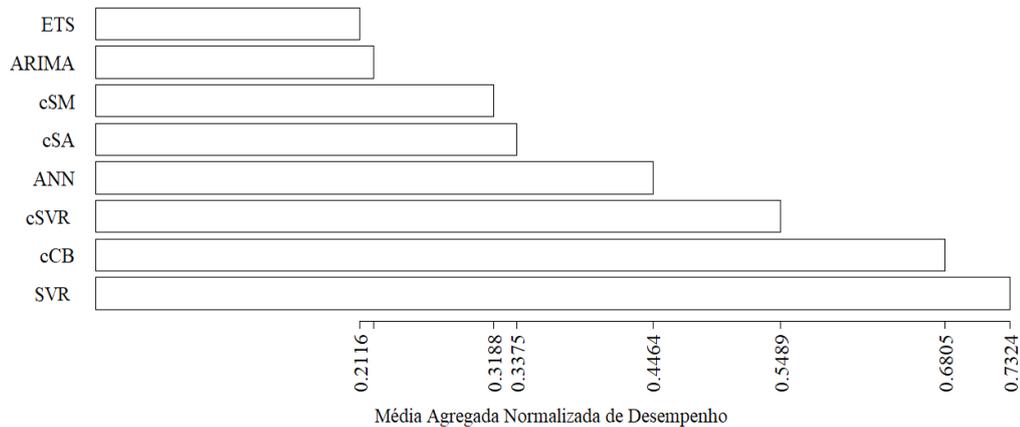
Tabela 25 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série ENE (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	51049.9	<b>32457.3</b>	32524.0	60710.2	37903.5	36041.8	48522.9	58592.2
MAPE	0.029	<b>0.023</b>	0.023	0.032	0.024	0.024	0.026	0.031
ARV	1.535	1.659	1.635	<b>1.314</b>	1.616	1.689	1.481	1.324
ID	0.502	0.413	<b>0.411</b>	0.506	0.450	0.445	0.467	0.495
Theil	1.427	<b>0.914</b>	0.916	1.694	1.062	1.011	1.368	1.634
WPOCID	<b>0.515</b>	0.529	0.529	0.588	0.544	0.544	0.559	0.544
Reg_Intercept	<b>3306.6</b>	3757.0	3760.8	3990.4	3670.5	3612.1	4423.1	3710.7
Reg_Slope	<b>0.514</b>	0.435	0.434	0.409	0.453	0.461	0.329	0.453
Reg_WR2	0.880	0.857	<b>0.855</b>	0.877	0.864	0.864	0.908	0.864

Fonte: Autor (2020).

Entre os preditores individuais, o ETS foi o melhor, enquanto que o cCB a pior combinação, como pode ser visto na Figura 44. Todos os modelos não foram melhores que a média regressa da série, com  $ARV > 1$ .

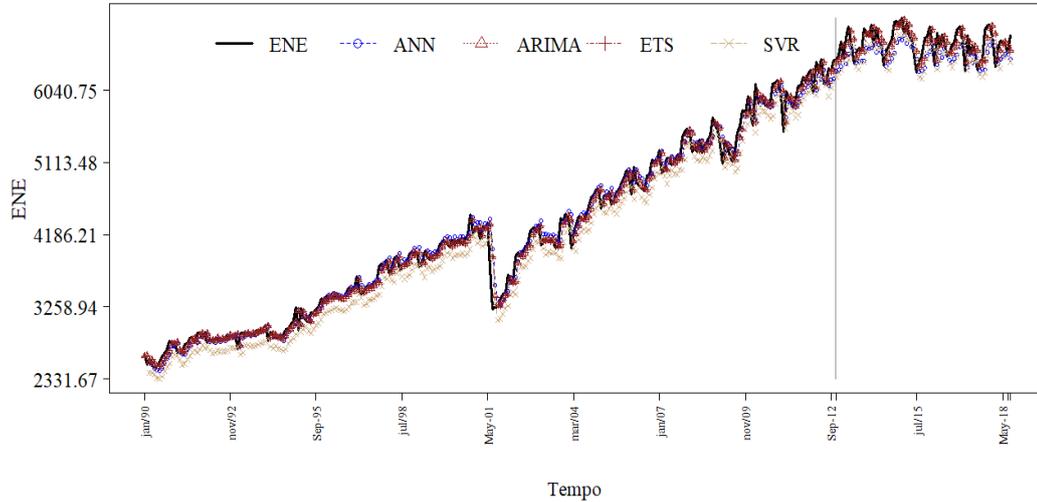
Figura 44 – Média Agregada de Performance dos modelos individuais e combinados para a série ENE.



Fonte: Autor (2020).

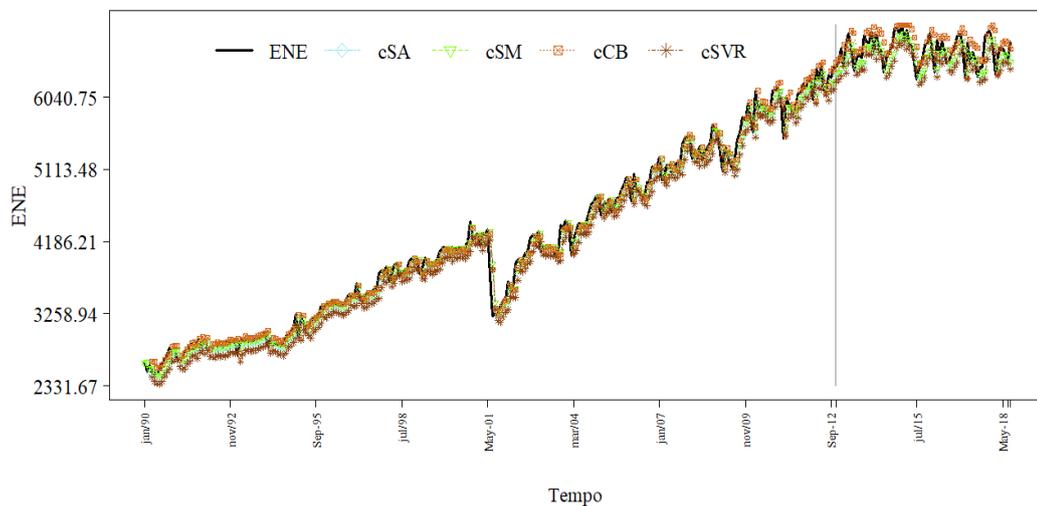
As Figuras 45 e 46 mostram os modelos individuais e combinados para a série ENE.

Figura 45 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série ENE (Fase de teste).



Fonte: Autor (2020).

Figura 46 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série ENE (Fase de teste).



Fonte: Autor (2020).

#### 4.3.4 Número de Casos de Caxumba em Nova York

A caxumba é uma doença viral infecciosa caracterizada pela inflamação da parótida e de outras glândulas salivares. Em Nova York essa doença é reportável, facilitando o acesso aos dados e sua análise (PATEL et al., 2017). Modelos de séries temporais ligados a

previsão de incidência de caxumba podem ser encontrados na literatura, como no trabalho de [Xu et al. \(2017\)](#) no qual foi usado o modelo SARIMA.

Na Tabela 26 estão caracterizados os modelos para a série MUC. O modelo ARIMA foi um MA(14) e o ETS foi uma Suavização Exponencial de Holt com erro multiplicativo e tendência multiplicativa amortecida. O modelo cCB apresentou marginais Normais e Laplace com conjunta Normal.

Tabela 26 – Descrição dos modelos individuais e combinados para a série MUC.

Formalismo	Modelo	Descrição
ANN	ANN(4, 1, 1)	$P_{AR} = 4, P_{ARS} = 0, LA = RPROP^+, FA = LHF$
ARIMA	ARIMA(0, 0, 14)	-
ETS	ETS(M,Md,N)	$\alpha = 1.000, \beta = 0.788, \phi = 0.803$
SVR	-	$P_{AR} = 1, P_{ARS} = 4, S = 33, cost = 58.3654, \epsilon = 0.9427,$ $kernel = LK, SV = 5$
cCB	-	$E_{ANN} \sim N(3 \times 10^{-4}, sd = 0.0175),$ $E_{ARIMA} \sim N(0, 0104, sd = 0.0161),$ $E_{ETS} \sim L(4, 6 \times 10^{-3}, sd = 9, 5 \times 10^{-3}),$ $E_{SVR} \sim N(4, 1 \times 10^{-3}, sd = 0.0149),$ $Cópula = Normal(\rho_1 = 0, 017) \rho_2 = 0, 3902,$ $\rho_3 = 0, 1018, \rho_4 = 0, 6189, \rho_5 = 0, 4144, \rho_6 = 0, 3321)$
cSVR	-	$cost = 56.2913, \epsilon = 0.8346, kernel = LK, SV = 3$

Fonte: Autor (2020).

O modelo cCB foi o melhor preditor entre todos, sendo o ARIMA o pior. A cSVR, com 3 vetores de suporte e função kernel polinomial, foi o pior preditor combinado. Todos os combinadores se destacaram sobre os modelos individuais como pode ser observado na Tabela 27.

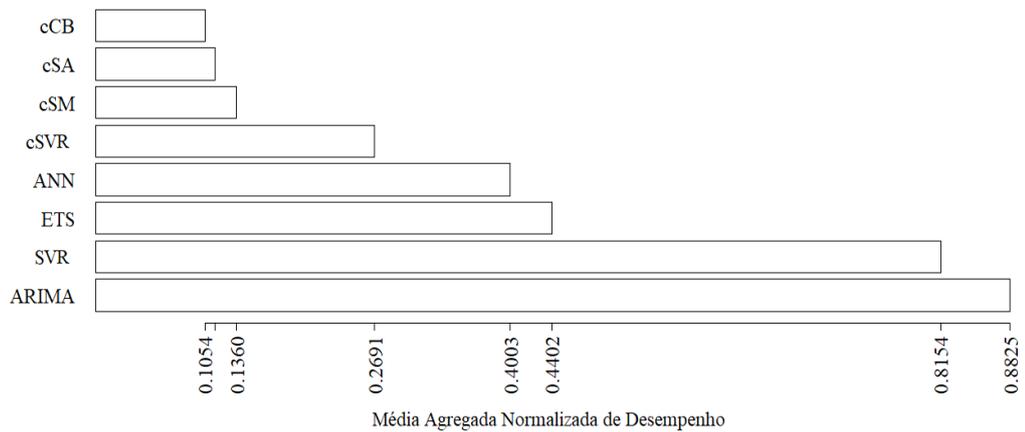
Tabela 27 – Performance dos modelos de previsão (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para a série MUC (fase de teste). Os melhores valores encontram-se em negrito.

Métrica	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
MSE	11047.5	20568.1	11874.0	18155.1	<b>8237.3</b>	8297.9	8642.3	11127.7
MAPE	0.375	0.488	0.268	0.507	<b>0.261</b>	0.271	0.289	0.372
ARV	0.326	0.319	0.223	0.290	0.178	<b>0.177</b>	0.241	0.287
ID	0.081	0.109	0.069	0.098	<b>0.051</b>	0.051	0.061	0.077
Theil	0.972	1.801	1.038	1.599	<b>0.725</b>	0.731	0.761	0.978
WPOCID	0.358	0.349	0.396	0.330	0.321	0.330	0.311	<b>0.302</b>
Reg_Intercept	-41.485	145.252	78.260	128.799	67.383	63.603	<b>-26.075</b>	-30.602
Reg_Slope	<b>0.966</b>	0.752	0.751	0.688	0.838	0.831	0.948	0.915
Reg_WR2	0.202	0.235	0.210	0.253	0.166	0.170	<b>0.163</b>	0.172

Fonte: Autor (2020).

A Figura 47 ilustra a média das métricas normalizadas que permite uma avaliação geral dos preditores.

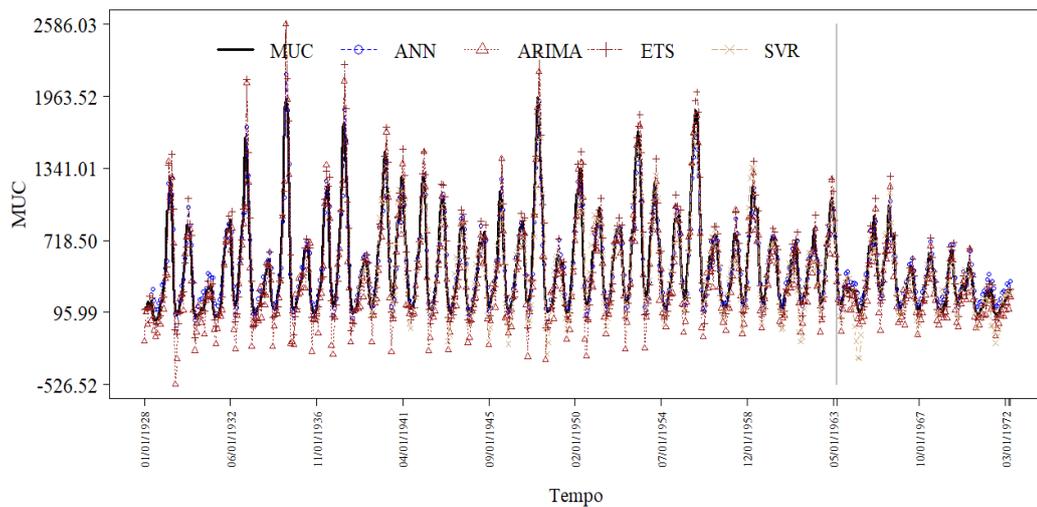
Figura 47 – Média Agregada de Performance dos modelos individuais e combinados para a série MUC.



Fonte: Autor (2020).

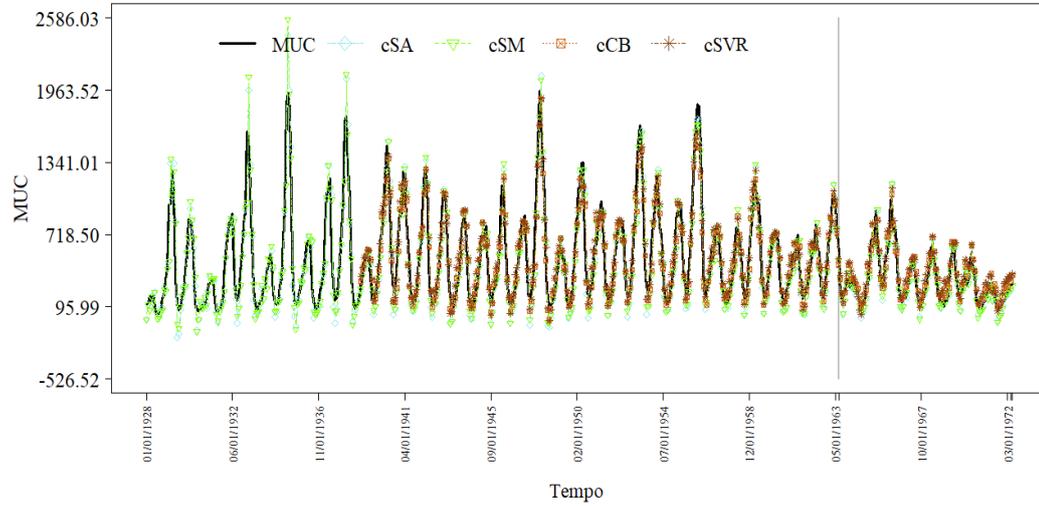
Para uma análise visual do desempenho dos preditores individuais e combinados pode-se recorrer às Figuras 48 e 49 a seguir.

Figura 48 – Modelos Individuais (ANN, ARIMA, ETS e SVR) para a Série MUC (Fase de teste).



Fonte: Autor (2020).

Figura 49 – Modelos Combinados (cSA, cSM, cCB e cSVR) para a Série MUC (Fase de teste).



Fonte: Autor (2020).

## 5 Discussão

Neste estudo foram usadas doze séries temporais para avaliar o desempenho dos modelos propostos. Foram gerados, ao todo, 96 modelos, sendo 48 individuais com as metodologias ANN, ARIMA, ETS e SVR e 48 combinados, cSA, cSM, cCB e cSVR. O método aqui empregado buscou analisar o potencial de cada preditor em variáveis ligadas ao desenvolvimento sustentável.

As Tabelas 28 e 29 a seguir mostram, respectivamente, os desempenhos dos modelos gerados sob a ótica da média agregada das métricas de desempenho normalizadas, para os conjuntos de teste e de treinamento. Sua análise permite notar que os modelos combinados não se sobressaíram diante dos modelos simples, vencendo em 5 das doze séries estudadas. Entretanto, o cCB foi o melhor modelo para 4 das doze séries, sendo o modelo que mais venceu de maneira geral.

O modelo de Redes Neurais, ANN, venceu em apenas uma das séries na fase de teste. Os modelos da classe ARIMA e ETS venceram em duas séries cada, (DEN e MVF) e (SP500 e ENE), respectivamente. O preditor individual SVR venceu os demais duas vezes (CO<sub>2</sub> e BBP). O combinador cSA venceu em apenas uma série (DEF).

Tabela 28 – Desempenho normalizado médio agregado dos modelos (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) ao prever as séries DEF, CO<sub>2</sub>, SUN, ONI, BBP, SP500, IPCA, PSP, MVF, MUC, ENE, DEN (Fase de teste). Os melhores valores encontram-se em negrito.

Série	Pior	Melhor	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
DEF	cCB	cSA	0.3405	<b>0.3377</b>	0.4163	0.5123	<b>0.3126</b>	0.3381	0.656	0.393
CO2	cSVR	SVR	0.0643	0.1358	0.2432	<b>0.0239</b>	0.06	<b>0.0472</b>	0.0705	0.9491
SUN	cCB	ANN	<b>0.0044</b>	0.0921	0.0231	0.5698	0.0709	<b>0.0225</b>	0.8251	0.8213
ONI	cSVR	cCB	0.0274	0.1442	<b>0.0302</b>	0.0484	0.0457	0.0307	<b>0.0173</b>	0.9175
BBP	cCB	SVR	0.3412	0.2953	0.2961	<b>0.1377</b>	0.1755	<b>0.1753</b>	0.7073	0.6317
SP500	ANN	ETS	0.8414	0.0408	<b>0.0186</b>	0.1466	0.126	<b>0.064</b>	0.1205	0.5287
IPCA	ARIMA	cCB	0.2289	0.7617	<b>0.2613</b>	0.2909	0.2345	0.2371	<b>0.2071</b>	0.5545
PSP	SVR	cCB	0.2857	<b>0.2754</b>	<b>0.2754</b>	0.9308	0.2772	0.2798	<b>0.0503</b>	0.5312
MVF	SVR	ARIMA	0.5793	<b>0.1726</b>	0.179	0.6916	0.3492	<b>0.2967</b>	0.5027	0.5653
MUC	ARIMA	cCB	0.4003	0.8825	<b>0.4402</b>	0.8154	0.1148	0.136	<b>0.1054</b>	0.2691
ENE	SVR	ETS	0.4464	0.2227	<b>0.2116</b>	0.7324	0.3375	<b>0.3188</b>	0.6805	0.5489
DEN	cSVR	ARIMA	0.4164	<b>0.0855</b>	0.0856	0.8779	0.276	<b>0.1716</b>	0.62	0.904

Fonte: Autor (2020).

Para as séries do eixo ambiental, os resultados da fase de teste foram bastante

distintos, sendo cSA melhor modelo para DEF, SVR para CO<sub>2</sub>, ANN para SUN e cCB para ONI. Em contrapartida, nesse segmento, os modelos cCB e cSVR foram os piores para duas séries cada. No eixo econômico, o modelo combinado via cópula venceu os demais em duas das quatro séries. E no eixo social, o ARIMA destacou-se em duas das quatro séries e o cCB e ETS, uma vez.

Dentre os modelos individuais, para as doze séries, os modelos ARIMA e ETS se mostraram melhores que os demais em 4 e 6 séries cada um, respectivamente. Já dentre os combinadores, o modelo cSM destacou-se em 7 séries, enquanto cCB destacou-se em 4 das doze séries.

O desempenho da combinação via SVR foi melhor no conjunto de treinamento. Entretanto, quando se deparou com dados desconhecidos, fase de teste, seu desempenho caiu consideravelmente. O modelo SVR individual venceu em 2 ocasiões na etapa de treinamento.

É necessário um ajuste mais refinado dos hiperparâmetros *cost* e  $\epsilon$  e os parâmetros da função kernel para melhorar o desempenho da SVR. Vale salientar que, em alguns casos, o modelo SVR consegue alcançar, no conjunto de treinamento, todos os valores "alvos" da série, com erro nulo. Entretanto, nesse caso, para o conjunto de treinamento, tem-se um pior resultado, no qual o modelo SVR retornou o valor médio da série.

O modelo ANN foi o melhor preditor individual na etapa de treinamento, vencendo em 7 das 12 séries, entretanto seu desempenho diante de dados novos caiu. Uma explicação razoável para isso pode ser o superajuste do modelo aos dados de treinamento, perdendo seu poder de generalização. Além disso, a não realização de uma etapa de validação pode ter contribuído.

Tabela 29 – Desempenho normalizado médio agregado dos modelos (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) ao prever as séries DEF, CO<sub>2</sub>, SUN, ONI, BBP, SP500, IPCA, PSP, MVF, MUC, ENE, DEN (Fase de treinamento). Os melhores valores encontram-se em negrito.

Série	Pior	Melhor	ANN	ARIMA	ETS	SVR	cSA	cSM	cCB	cSVR
DEF	cCB	ANN	<b>0.1281</b>	0.2816	0.38	0.2623	0.1583	0.2165	0.7499	<b>0.1468</b>
CO2	ETS	cCB	<b>0.0672</b>	0.2545	0.7778	0.3801	0.1808	0.1832	<b>0.0436</b>	0.2005
SUN	SVR	cSVR	<b>0.4174</b>	0.6395	0.4288	0.6856	0.4319	0.4307	0.1977	<b>0</b>
ONI	cSVR	ANN	<b>0</b>	0.1995	0.0289	0.0515	0.0561	0.0324	<b>0.0262</b>	0.9541
BBP	cCB	cSM	<b>0.069</b>	0.102	0.0893	0.5146	0.0666	<b>0.0515</b>	0.7543	0.4741
SP500	cSVR	ETS	0.5477	0.1132	<b>0.099</b>	0.2612	0.1209	<b>0.1144</b>	0.3213	0.8435
IPCA	ARIMA	SVR	0.2706	0.819	0.3228	<b>0.0073</b>	0.2216	0.2538	<b>0.1323</b>	0.3079
PSP	SVR	ANN	<b>0.0251</b>	0.1786	0.1784	0.9216	0.2707	<b>0.1061</b>	0.2238	0.5467
MVF	cCB	cSA	0.4307	0.489	0.5362	<b>0.3511</b>	<b>0.1698</b>	0.1905	0.802	0.5989
MUC	ARIMA	cCB	<b>0.1806</b>	0.9919	0.6912	0.4136	0.3618	0.3334	<b>0.0283</b>	0.0688
ENE	SVR	ETS	0.1881	0.1284	<b>0.12</b>	0.7269	0.1822	<b>0.1735</b>	0.295	0.4957
DEN	ETS	SVR	0.3186	0.9058	0.906	<b>0.0426</b>	0.3297	0.4529	<b>0.1097</b>	0.0735

Fonte: Autor (2020).

Há de ser considerada também a parcimônia dos modelos. O princípio da parcimônia estabelece que quando há concorrência entre modelos equivalentes, deve-se optar pelo mais simples. No âmbito do aprendizado de máquina, este princípio considera a aplicabilidade do modelo a diferentes bases de dados, o custo computacional e o tempo gasto no processo de modelagem (PAULA, 2016).

Outro ponto a ser discutido é o tempo de treinamento dos modelos. A Tabela 30 mostra o tempo gasto para o treinamento de cada modelo, bem como o tempo médio por série, por modelo e geral. Em termos gerais, o tempo médio de modelagem das alternativas adotadas foi de 6,630 segundos. Dentre os formalismos de modelagem individual, o que consumiu menor tempo, em média, foi o ETS; cerca de 0,187 segundos. Já dentre os combinadores paramétricos, cCB foi o mais rápido, envolvendo em média 0,427 segundos. Por outro lado, em relação às séries temporais, ENE foi a que demandou mais tempo de modelagem; em média 11,7093 segundos. Em média, o modelo que demandou mais tempo na etapa de treinamento foi o ANN, seguido do SVR e cSVR.

Tabela 30 – Tempo de treinamento dos modelos (ANN, ARIMA, ETS, SVR, cSA, cSM, cCB, cSVR) para as séries DEF, CO<sub>2</sub>, SUN, ONI, BBP, SP500, IPCA, PSP, MVF, MUC, ENE, DEN (em segundos). Os melhores valores encontram-se em negrito.

Série	ANN	ARIMA	ETS	SVR	cCB	cSVR	Média
DEF	10.0164	<b>0.0232</b>	0.0339	10.611	0.1713	5.6324	4.4147
CO2	32.6734	<b>0.0987</b>	0.2434	13.340	1.0023	11.5319	9.81495
SUN	14.6194	10.0335	0.758	1.102	<b>0.3394</b>	2.9379	4.965033
ONI	1.1494	17.5301	<b>0.3291</b>	29.224	1.0711	14.3365	10.6067
BBP	10.044	<b>0.012</b>	0.0529	10.073	0.1646	10.3501	5.1161
SP500	13.0493	<b>0.0199</b>	0.2294	10.232	1.0452	11.5893	6.027517
IPCA	15.6006	0.4468	<b>0.0957</b>	10.390	0.6144	6.4195	5.5945
PSP	14.9596	<b>0.0239</b>	0.0878	10.026	0.373	10.3633	5.972267
MVF	10.0037	<b>0.0279</b>	0.0389	10.625	0.2045	1.6505	<b>3.758417</b>
MUC	10.1273	3.8228	<b>0.2394</b>	11.319	0.7181	10.9818	6.2014
ENE	47.7248	<b>0.018</b>	0.0938	10.55	0.3929	11.4763	11.7093
DEN	10.1457	<b>0.007</b>	0.0389	11.774	0.2234	10.0518	5.373467
Média	15.8428	2.671983	<b>0.1867667</b>	11.60572	0.5266833	8.943442	6.629529

Fonte: Autor (2020).

Vale destacar que os modelos de SVR tenderam a envolver a maior quantidade de parâmetros a otimizar, uma vez que as Funções Kernel possuem parâmetros que também foram levados em consideração para a otimização. Com isso, o desempenho das combinações pode ter sido prejudicado pelo baixo desempenho dos modelos individuais ANN e SVR na etapa de teste.

## 6 Conclusão

Nessa dissertação foi abordado o problema de previsão de séries temporais sob o olhar dos preditores combinados. Com o intuito de avaliar estatisticamente o desempenho das combinações, dedicou-se a aplicações em séries reais na temática do desenvolvimento sustentável, valendo-se de quatro tipos de metodologias combinadas (cSA, cSM, cCB e cSVR). Para os modelos simples, foram adotadas as modelagem ANN, ARIMA, ETS e SVR.

Os resultados das métricas de desempenho para as aplicações aqui realizadas mostram que os preditores combinados não apresentaram melhor desempenho que os individuais, todavia, a combinação via cópula teve o melhor resultado entre todos os preditores. As combinações podem ter sido influenciadas negativamente pelo baixo desempenho do SVR e ANN individuais. A combinação via SVR mostrou desempenho inferior ao combinador baseado em cópulas, sendo menos aconselhável diante destes para os casos abordados.

Quanto aos modelos individuais, os preditores ARIMA e ETS foram os melhores, enquanto ANN e SVR foram os piores. Uma possível justificativa para o desempenho precário de SVR e ANN reside no possível superajuste às séries de treinamento, mesmo sendo baseados em um critério de informação naturalmente compromissado com a parcimônia, o BIC, que pareceu ser muito severo com os modelos ANN e SVR.

### 6.1 Limitações da Pesquisa

- Não adoção de um critério de informação mais adequado para SVR e ANN no problema de séries temporais, considerando suas particularidades, uma vez que envolvem o uso de maior número de parâmetros a serem otimizados;
- O envolvimento de séries temporais relativamente pequenas;
- Entendimento ainda inicial quanto às nuances das séries temporais voltadas aos eixos básicos do desenvolvimento sustentável (ambiental, social, econômico).

## 6.2 Trabalhos Futuros

- Estudar a parcimônia dos modelos de *Support Vector Regression* e Redes Neurais individuais e combinados;
- Trabalhar com outros formalismos individuais, como as metodologia *Fuzzy* e GARCH, e combinados, como os baseados em ANN e Mínima Variância e seus desempenhos em séries temporais relevantes.
- Estudar alternativas de algoritmos para otimização de Redes Neurais e *Support Vector Regression* que proporcionem um balanço dinâmico entre os processos de *exploration* e *exploitation*.
- Explorar o potencial de outras distribuições marginais para cópulas, a exemplo da Cauchy Truncada.

## Referências Bibliográficas

- ABREU, D. R. de O. M.; SOUZA, E. M. de; MATHIAS, T. A. de F. Impacto do código de trânsito brasileiro e da lei seca na mortalidade por acidentes de trânsito. **Cadernos de Saúde Pública**, SciELO Public Health, v. 34, p. 117 –122, 2018.
- ANP. **Agência Nacional do Petróleo**. 2018. Disponível em: <<http://www.anp.gov.br/>>.
- ANP. **Agência Nacional do Petróleo, Gás Natural e Biocombustíveis**. 2019. Disponível em: <<http://www.anp.gov.br/biocombustiveis>>.
- ASSEMBLY, G. sustainable development goals. **SDGs), Transforming our world: the**, v. 2030, 2015.
- ASSIS, T. F. O. **Cópulas para a combinação de modelos de séries temporais**. Tese (Doutorado) — Universidade Federal Rural de Pernambuco, 2017.
- ATREDINGVIEW. **Ações da Petrobrás**. 2018. Disponível em: <<https://br.tradingview.com/symbols/BMFBOVESPAPETR4/>>.
- AZZALINI, A. A class of distributions which includes the normal ones. **Scandinavian journal of statistics**, JSTOR, p. 171–178, 1985.
- BAQUERO, O. S.; SANTANA, L. M. R.; CHIARAVALLOTI-NETO, F. Dengue forecasting in são paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. **PloS one**, Public Library of Science, v. 13, n. 4, 2018.
- BARAK, S.; SADEGH, S. S. Forecasting energy consumption using ensemble arima–anfis hybrid algorithm. **International Journal of Electrical Power & Energy Systems**, Elsevier, v. 82, p. 92–104, 2016.
- BARROS, A. C. et al. **Análise de séries temporais em R: um curso introdutório**. Rio de Janeiro, BR: Elsevier: FGV IBRE, 2018.
- BATES, J. M.; GRANGER, C. W. The combination of forecasts. **Journal of the Operational Research Society**, Taylor & Francis, v. 20, n. 4, p. 451–468, 1969.
- BAUER, R. A. **Social Indicators**. [S.l.]: Massachusetts Inst. of Technology, 1966.
- BEALE, R.; JACKSON, T. **Neural Computing: An Introduction**. New York , USA: Adam Hilger, 1991.
- BHAT, C. R.; ELURU, N. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. **Transportation Research Part B: Methodological**, Elsevier, v. 43, n. 7, p. 749–765, 2009.

- BHATTI, M. I.; DO, H. Q. Recent development in copula and its applications to the energy, forestry and environmental sciences. **International journal of hydrogen energy**, v. 44, p. 453–473, 2019.
- BOONTOME, P.; THERDYOTHIN, A.; CHONTANAWAT, J. Forecasting carbon dioxide emission and sustainable economy: Evidence and policy responses. **International Journal of Energy Economics and Policy**, Econjournals, v. 9, n. 5, p. 55, 2019.
- BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. p. 144–152, 1992.
- BOULANGER, P.-M. Sustainable development indicators: a scientific challenge, a democratic issue. **Sapiens**, v. 1, n. 1, p. 45–59, 2008.
- BOX, G. E.; JENKINS, P.; REINSEL, G. M. G. C. **Time series analysis: Forecasting and control**. [S.l.]: John Wiley & Sons, 2008.
- BOZIC, M.; STOJANOVIC, M. Application of svm methods for mid-term load forecasting. **Serbian Journal of Electrical Engineering**, v. 8, p. 73–83, 2011. Disponível em: <[http://www.journal.ftn.kg.ac.rs/Vol\\_8-1/08-Bozic-Stojanovic.pdf](http://www.journal.ftn.kg.ac.rs/Vol_8-1/08-Bozic-Stojanovic.pdf)>.
- BRAGA, A. de P.; CARVALHO, A. P. de L. F. de; LUDEMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. Rio de Janeiro, BR: LTC, 2007.
- CARVALHO, S.; MAGALHÃES, M. d. A. F. M.; MEDRONHO, R. d. A. Análise da distribuição espacial de casos da dengue no município do rio de janeiro, 2011 e 2012. **Revista de Saúde Pública**, SciELO Public Health, v. 51, p. 79, 2017.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. [S.l.]: Duxbury Pacific Grove, CA, 2002. v. 2.
- CHATFIELD, C. **Time-series forecasting**. Washington, USA: Chapman & Hall/CRC, 2000.
- CHEN, N. et al. **Support Vector Machine in chemistry**. [S.l.]: World Scientific, 2004.
- CHERUBINI, U. et al. **Dynamic Copula Methods in Finance**. [S.l.]: John Wiley Sons, 2012.
- COMPUTING, R. F. for S. **A language and environment for statistical computing**. 2015. Disponível em: <<https://www.R-project.org/>>
- CROWTHER, D.; SEIFI, S.; MOYEEN, A. Responsibility and governance in achieving sustainability. In: **The Goals of Sustainable Development**. Melbourne, Australia,: [s.n.], 2018. p. 1–18.
- DATAMARKET. **Monthly reports of mumps cases in New York City**. 2018. Disponível em: <<https://datamarket.com/data/list/?q=provider:tsdl>>.
- DETER. **Detecção de desmatamento em tempo real**. 2018. Disponível em: <<http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/deter>>.

- DETRAN. **Acidentes com vítimas fatais**. 2018. Disponível em: <<http://portal.detran.ce.gov.br/index.php/estatisticas>>.
- DOMINGOS, S. d. O.; OLIVEIRA, J. F. de; NETO, P. S. de M. An intelligent hybridization of arima with machine learning models for time series forecasting. **Knowledge-Based Systems**, Elsevier, v. 175, p. 72–86, 2019.
- DONG, Z. et al. A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance. **Energy**, v. 82, p. 570–577, 2015.
- FEARNSIDE, P. M. Desmatamento na amazônia brasileira: história, índices e consequências. **Megadiversidade**, v. 1, n. 1, p. 113–123, 2005.
- FEARNSIDE, P. M. Desmatamento na amazônia: dinâmica, impactos e controle. v. 36, n. 3, p. 395 – 400, 2006.
- FEARNSIDE, P. M. Como sempre, os negócios: o ressurgimento do desmatamento na amazônia brasileira. **Yale Environment**, v. 360, p. 18, 2017.
- FIRMINO, P. R. A.; NETO, P. S. de M.; FERREIRA, T. A. Correcting and combining time series forecasters. **Neural Networks**, Elsevier, v. 50, p. 1–11, 2014.
- FIRMINO, P. R. A.; NETO, P. S. de M.; FERREIRA, T. A. Error modeling approach to improve time series forecasters. **Neurocomputing**, Elsevier, v. 153, p. 242–254, 2015.
- FIRMINO, P. R. A. e. a. copulas-based ensemble of artificial neural networks for forecasting real world time series. **International Joint Conference on Neural Networks**, 2016. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/7727732>>.
- FRITSCH, S.; GUENTHER, F.; WRIGHT, M. N. **neuralnet: Training of Neural Networks**. [S.l.], 2019. R package version 1.44.2. Disponível em: <<https://CRAN.R-project.org/package=neuralnet>>.
- GARDNER, E. S. Exponential smoothing: The state of the art,. **Journal of Forecasting**, v. 4, n. 1, p. 1–28, 1985.
- GRI. Sustainability and reporting trends in 2025: Preparing for the future. In: \_\_\_\_\_. [S.l.: s.n.], 2015.
- GUTIERREZ, L. Impacts of el niño-southern oscillation on the wheat market: A global dynamic analysis. **PloS one**, Public Library of Science, v. 12, n. 6, 2017.
- HARRIS, J. M. Basic principles of sustainable development. In: \_\_\_\_\_. [S.l.: s.n.], 2001. p. 21–40.
- HAYKIN, S. **Redes Neurais: princípios e prática**. Porto Alegre, BR: Bookman, 2001.
- HAYKIN, S. **Neural Networks and Learning Machines**. New Jersey, EUA: Pearson, 2009.

- HOFERT, M.; MÄCHLER, M. Nested archimedean copulas meet R: The nacopula package. **Journal of Statistical Software**, v. 39, n. 9, p. 1–20, 2011. Disponível em: <<http://www.jstatsoft.org/v39/i09/>>.
- HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the national academy of sciences**, National Acad Sciences, v. 79, n. 8, p. 2554–2558, 1982.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, v. 26, n. 3, p. 1–22, 2008. Disponível em: <<http://www.jstatsoft.org/article/view/v027i03>>.
- HYNDMAN, R. J.; KOEHLER, A. B. **Forecasting with exponential smoothing**. New York , USA: Springer, 2008.
- HYNDMAN, R. J. et al. A state space framework for automatic forecasting using exponential smoothing methods. **International Journal of Forecasting**, v. 18, n. 3, p. 439–454, 2002.
- IBGE. **Série histórica anual IPCA**. 2018. Disponível em: <[https://ww2.ibge.gov.br/english/estatistica/indicadores/precos/inpc\\_ipca/defaultinpc.shtm](https://ww2.ibge.gov.br/english/estatistica/indicadores/precos/inpc_ipca/defaultinpc.shtm)>.
- IBGE. **Instituto Brasileiro de Geografia e Estatística**. 2019. Disponível em: <<https://www.ibge.gov.br/explica/inflacao.php>>.
- IGEL, C.; HÜSKEN, M. Empirical evaluation of the improved rprop learning algorithms. **Neurocomputing**, v. 50, p. 105–123, 2003.
- IPEA. **Consumo mensal de energia elétrica no nordeste brasileiro**. 2018. Disponível em: <<http://www.ipeadata.gov.br/>>.
- JAGANATHAN, S.; PRAKASH, P. A combination-based forecasting method for the m4-competition. **International Journal of Forecasting**, Elsevier, v. 36, n. 1, p. 98–104, 2020.
- JOE, H.; XU, J. J. The estimation method of inference functions for margins for multivariate models. **Journal of Computing in Civil Engineering**, n. 166, 1996.
- JOSHI, D. K.; HUGHES, B. B.; SISK, T. D. Improving governance for the post-2015 sustainable development goals: scenario forecasting the next 50 years. **World Development**, Elsevier, v. 70, p. 286–302, 2015.
- KANG, F.; LI, J. Artificial bee colony algorithm optimized support vector regression for system reliability analysis of slopes. **Journal of Computing in Civil Engineering**, v. 30, p. 3–14, 2016.
- KEEBLE, B. R. The brundtland report:‘our common future’. **Medicine and War**, Taylor & Francis, v. 4, n. 1, p. 17–25, 1988.
- KIM, K. Financial time series forecasting using support vector machines. **Neurocomputing**, v. 55, p. 307 – 319, 2003. Disponível em: <<http://lcsd2.svms.org/finance/Kim2003.pdf>>.

- KIM, S.; KIM, H. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, Elsevier, v. 32, n. 3, p. 669–679, 2016.
- KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. **science**, American Association for the Advancement of Science, v. 220, n. 4598, p. 671–680, 1983.
- KOTZ, S.; KOZUBOWSKI, T.; PODGORSKI, K. **The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance**. [S.l.]: Springer Science & Business Media, 2012.
- KUROWICKA, D.; JOE, H. **Dependence Modeling Vine Copula Handbook**. [S.l.]: World Scientific, 2011.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.
- LERNER, A. Y.; VAPNIK, V. N. Recognition of patterns with help of generalized portraits. **Avtomat. i Telemekh**, v. 24, n. 6, p. 774–780, 1963.
- LI, F. et al. Modelling of a post-combustion co2 capture process using deep belief network. **Applied Thermal Engineering**, Elsevier, v. 130, p. 997–1003, 2018.
- LI, G.; WANG, S. Sunspots time-series prediction based on complementary ensemble empirical mode decomposition and wavelet neural network. **Mathematical Problems in Engineering**, Hindawi, v. 2017, 2017.
- LIU, C. et al. Forecasting s&p 500 stock index using statistical learning models. **Open journal of statistics**, Scientific Research Publishing, v. 6, n. 6, p. 1067–1075, 2016.
- LORENA, A. C.; CARVALHO, A. C. de. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43–67, 2007.
- MAGALHÃES, M. N. **probabilidade e Variáveis Aleatórias**. [S.l.]: EdUSP, 2015.
- MAGAZZINO, C. The relationship between real gdp, co2 emissions, and energy use in the gcc countries: A time series approach. **Cogent Economics & Finance**, Taylor & Francis, v. 4, n. 1, p. 1152729, 2016.
- MCCLELLAND, J. L.; RUMELHART, D. E. **Parallel Distributed Processing: Explorations in the Microstructure of Cognition**. Cambridge: MIT Press, 1986.
- MCCORMICK, G. P. Communications to the editor—exponential forecasting: Some new variations. **Management Science**, v. 15, n. 5, p. 311–320, 1969.
- MENG, J. et al. Complexity-based approach for el niño magnitude forecasting before the spring predictability barrier. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 117, n. 1, p. 177–183, 2020.

- MEYER, D. et al. **e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien**. [S.l.], 2019. R package version 1.7-2. Disponível em: <<https://CRAN.R-project.org/package=e1071>>.
- MITCHAM, C. The concept of sustainable development: its origins and ambivalence. **Technology In Somy**, v. 17, n. 3, p. 311–326., 1995.
- MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to time series analysis and forecasting**. New Jersey, USA: John Wiley & Sons, 2015.
- MORETTIN, P. A.; TOLOI, C. M. **Análise de Séries Temporais**. [S.l.]: Blucher, 2018.
- MUSCHETT, D. **Principles of Sustainable Development**. [S.l.]: CRC Press, 1996.
- NEATH, A. A.; CAVANAUGH, J. E. The bayesian information criterion: background, derivation, and applications. **Wiley Interdisciplinary Reviews: Computational Statistics**, Wiley Online Library, v. 4, n. 2, p. 199–203, 2012.
- NELSEN, R. B. **An introduction to copulas**. Portland, USA: Springer, 2013.
- NETO, M. C. A.; CALVALCANTI, G. D.; REN, T. I. Financial time series prediction using exogenous series and combined neural networks. In: IEEE. **2009 International Joint Conference on Neural Networks**. [S.l.], 2009. p. 149–156.
- NETO, P. S. de M. et al. A temporal-window framework for modeling and forecasting time series. **Knowledge-Based Systems**, Elsevier, p. 105476, 2020.
- NGUYEN, P.; HALEM, M. Prediction of co2 flux using long short term memory (lstm) recurrent neural networks with data from flux towers and oco-2 remote sensing. In: **AGU Fall Meeting Abstracts**. [S.l.: s.n.], 2018.
- NIETO, P. G. et al. Hybrid pso–svm-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. **Reliability Engineering & System Safety**, Elsevier, v. 138, p. 219–231, 2015.
- NOAA. **Administração nacional oceânica e atmosférica: episódios quentes e frios por temporada**. 2018. Disponível em: <[http://origin.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ONI\\_v5.php](http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php)>.
- NOAA. **Administração nacional oceânica e atmosférica: tendências do dióxido de carbono atmosférico**. 2018. Disponível em: <<https://www.esrl.noaa.gov/gmd/ccgg/trends/>>.
- NYONI, T.; BONGA, W. G. Prediction of co2 emissions in india using arima models. **DRJ-Journal of Economics & Finance**, v. 4, n. 2, p. 01–10, 2019.
- OKASHA, M. K. Using support vector machines in financial time series forecasting. **International Journal of Statistics and Applications**, v. 4, p. 28–39, 2014. Disponível em: <<http://article.sapub.org/10.5923.j.statistics.20140401.03.html>>.

OLIVEIRA, F. A. de; NOBRE, C. N.; ZARATE, L. E. Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index—case study of petr4, petrobras, brazil. **Expert Systems with Applications**, Elsevier, v. 40, n. 18, p. 7596–7606, 2013.

OLIVEIRA, R. et al. Copulas-based time series combined forecasters. **Information Sciences**, v. 376, 2016.

OLIVEIRA, R. T. A. **Modelagem e simulação computacional da combinação de preditores de séries temporais por meio de cópulas**. Dissertação (Mestrado) — Universidade Federal Rural de Pernambuco, 2014.

OLIVEIRA, R. T. A. e. a. Copulas-based time series combined forecasters. **Information Sciences**, v. 376, p. 110–124, 2017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0020025516312397>>.

OLIVEIRA, R. T. A. e. a. Aggregation of time series forecasts via cacoulllos copula. **International Joint Conference on Neural Networks**, 2018. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8489098>>.

OZTURK, S.; OZTURK, F. Forecasting energy consumption of turkey by arima model. **Journal of Asian Scientific Research**, Asian Economic and Social Society, v. 8, n. 2, p. 52, 2018.

PATEL, L. N. et al. Mumps outbreak among a highly vaccinated university community—new york city, january–april 2014. **Clinical Infectious Diseases**, Oxford University Press US, v. 64, n. 4, p. 408–412, 2017.

PATTON, A. J. A review of copula models for economic time series. **Journal of Multivariate Analysis**, Elsevier, v. 110, p. 4–18, 2012.

PAULA, R. O. d. Máquinas de suporte vetorial como instrumento de priorização de investimentos aplicado ao programa de investimento em logística–aeroportos. 2016.

PETROBRAS. 2019. Disponível em: <<https://petrobras.com.br/pt/>>.

PETROPOULOS, F.; SVETUNKOV, I. A simple combination of univariate models. **International journal of forecasting**, Elsevier, v. 36, n. 1, p. 110–115, 2020.

REN, Y.; BAI, G. Determination of optimal svm parameters by using ga/psa. **Journal of Computers**, v. 5, n. 8, p. 1160–1168, 2010.

RENARD, B.; LANG, M. Use of a gaussian copula for multivariate extreme value analysis: some case studies in hydrology. **Advances in Water Resources**, Elsevier, v. 30, n. 4, p. 897–912, 2007.

RIEDMILLER, M. Advanced supervised learning in multi-layer perceptrons - from backpropagation to adaptive learning algorithms. **Computer Standards Interfaces**, v. 16, n. 3, p. 265–278, 1994.

- RIEDMILLER, M.; BRAUN, H. A direct adaptive method for faster backpropagation learning the rprop algorithm. **Proceedings of the IEEE international conference on neural networks**, p. 586–591, 1993.
- ROUT, A. K. et al. Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. **Journal of King Saud University-Computer and Information Sciences**, Elsevier, v. 29, n. 4, p. 536–552, 2017.
- SADAEI, H. J. et al. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. **Energy**, Elsevier, v. 175, p. 365–377, 2019.
- SAPANKEVYCH, N. I.; SANKAR, R. Time series prediction using support vector machines: a survey. **IEEE Computational Intelligence Magazine**, IEEE, v. 4, n. 2, p. 24–38, 2009.
- SCHWARZ, G. Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.
- SKLAR, A. Distribution functions of n dimensions and margins. **Institute of Statistics at the University of Paris**, p. 229–231, 1959.
- SMOLA, A. J.; SCHOLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, p. 199–222, 2004.
- SMS-RJ. **Número mensal reportado de casos de dengue na cidade do Rio de Janeiro**. 2018. Disponível em: <<http://www.rio.rj.gov.br/web/sms/dengue>>.
- STOLERMAN, L. M.; MAIA, P. D.; KUTZ, J. N. Forecasting dengue fever in brazil: An assessment of climate conditions. **PloS one**, Public Library of Science, v. 14, n. 8, 2019.
- SU, J. et al. Ga-based support vector machine model for the prediction of monthly reservoir storage. **Journal of Hydrologic Engineering**, American Society of Civil Engineers, v. 19, n. 7, p. 1430–1437, 2013.
- SUN, Y.-Y. et al. El niño–southern oscillation effect on quasi-biennial oscillations of temperature diurnal tides in the mesosphere and lower thermosphere. **Earth, Planets and Space**, SpringerOpen, v. 70, n. 1, p. 1–10, 2018.
- TAYLOR, J. W. Exponential smoothing with a damped multiplicative trend. **International journal of Forecasting**, v. 19, n. 4, p. 715–725, 2003.
- TONG, H. **Non-linear time series: a dynamical system approach**. [S.l.: s.n.], 1990.
- TRIVEDI, P. K.; ZIMMER, D. M. **Copula Modeling: An Introduction for Practitioners**. [S.l.]: Foundations and Trends in Econometrics, 2011.
- VAPNIK, V. **The nature of statistical learning theory**. New York, USA: Springer, 1995.

- VAPNIK, V. **The nature of statistical learning theory**. New York, USA: Springer, 2000.
- VESCO, A.; FERRERO, F. **Handbook of research on social, economic, and environmental sustainability in the development of smart cities**. [S.l.]: IGI Global, 2015.
- WANG, X. et al. Real estate price forecasting based on svm optimized by pso. **Optik-International Journal for Light and Electron Optics**, Elsevier, v. 125, n. 3, p. 1439–1443, 2014.
- WHITE, S. R. Concepts of scale in simulated annealing. In: AIP. **AIP Conference Proceedings**. [S.l.], 1984. v. 122, n. 1, p. 261–270.
- WILLMOTT, C. J. Some comments on the evaluation of model performance. **Bulletin of the American Meteorological Society**, American Meteorological Society, v. 63, n. 11, p. 1309–1313, 1982.
- WU, M. et al. Is urban development ecologically sustainable? ecological footprint analysis and prediction based on a modified artificial neural network model: A case study of tianjin in china. **Journal of Cleaner Production**, Elsevier, v. 237, p. 117795, 2019.
- XAVIER, E. F. M. et al. Modelagem e previsão da produção brasileira de biodiesel. **Gaia Scientia**, v. 12, n. 3, p. 155–167, 2018.
- XIANG, Y. et al. Generalized simulated annealing for efficient global optimization: the GenSA package for R. **The R Journal Volume 5/1, June 2013**, 2013. Disponível em: <<https://journal.r-project.org/archive/2013/RJ-2013-002/index.html>>.
- XIE, Y.; LOU, Y. Hydrological time series prediction by arima-svr combined model based on wavelet transform. In: **Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence**. [S.l.: s.n.], 2019. p. 243–247.
- XU, Q. et al. Forecasting the incidence of mumps in zibo city based on a sarima model. **International journal of environmental research and public health**, Multidisciplinary Digital Publishing Institute, v. 14, n. 8, p. 925, 2017.
- YAHOOFINANCE. **Ações Goog**. 2018. Disponível em: <<https://finance.yahoo.com/quote/GOOG/>>.
- YAN, J. Enjoy the joy of copulas: with a package copula. **Journal of Statistical Software**, v. 21, n. 4, 2007.
- ZHANG, F. et al. Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique. **Energy and Buildings**, Elsevier, v. 126, p. 94–103, 2016.