

UNIVERSIDADE FEDERAL DO PARANÁ

GABRIEL TRIERWEILER RIBEIRO



PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA DE CURTO PRAZO
UTILIZANDO ABORDAGENS DE COMITÊS DE *WAVENETS*

CURITIBA

2017

GABRIEL TRIERWEILER RIBEIRO

PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA DE CURTO PRAZO
UTILIZANDO ABORDAGENS DE COMITÊS DE *WAVENETS*

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, Área de Concentração Sistema Eletrônicos – Controle e Automação, Departamento de Engenharia Elétrica, Setor de Tecnologia, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Leandro dos Santos Coelho

CURITIBA

2017



UNIVERSIDADE FEDERAL DO PARANÁ
Programa de Pós-Graduação em Engenharia Elétrica, PPGE
Setor de Tecnologia

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ENGENHARIA ELÉTRICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de GABRIEL TRIERWEILER RIBEIRO intitulada: **Previsão de demanda de energia elétrica de curto prazo utilizando abordagens de comitês de wavenets**, após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua aprovação.

Curitiba, 12 de Abril de 2017.

LEANDRO DOS SANTOS COELHO
Presidente da Banca Examinadora (UFPR)

ALEXANDRE RASI AOKI
Avaliador Interno (UFPR)

JOÃO ALBERTO FABRO
Avaliador externo (UTFPR)

AGRADECIMENTOS

A Deus por ter me dado saúde, sabedoria e serenidade para superar todas as dificuldades.

Ao meu orientador professor Leandro dos Santos Coelho por seu incentivo, orientação, ensinamentos e todo o seu auxílio no desenvolvimento deste trabalho.

À minha família, em especial a meu pai Claudiney Armindo Ribeiro Neto que sempre se esforçou para me dar a melhor educação possível e condições de cursar a graduação.

À UFPR e a todos que direta ou indiretamente contribuíram de alguma forma para que este objetivo fosse alcançado.

E principalmente a minha querida esposa Danielle, por reconhecer a importância do tempo e recursos dedicados a este trabalho como parte do processo de busca da nossa felicidade.

“Não acredite em algo simplesmente porque ouviu. Não acredite em algo simplesmente porque todos falam a respeito. Não acredite em algo simplesmente porque está escrito em seus livros religiosos. Não acredite em algo só porque seus professores e mestres dizem que é verdade. Não acredite em tradições só porque foram passadas de geração em geração. Mas depois de muita análise e observação, se você vê que algo concorda com a razão, e que conduz ao bem e benefício de todos, aceite-o e viva-o”

RESUMO

A energia elétrica faz parte de um mercado que envolve agentes de geração, transmissão, distribuição e consumo que desejam maximizar seus lucros e minimizar suas despesas. Para isso precisam de um planejamento que tenha como base uma previsão de demanda precisa, já que um cenário pessimista pode levar ao despacho de mais geradores do que o necessário, reserva excessiva de matéria prima e aumento do custo de operação, e por outro lado um cenário otimista pode colocar o sistema elétrico em risco ou exigir a compra de energia no mercado livre a um preço alto. Por isso, a previsão de demanda tem sido empregada em áreas como o agendamento ótimo de geradores, planejamento da manutenção, planejamento da reserva hídrica, compreensão do padrão de consumo, planejamento da expansão e previsão de preços e ajuste de tarifas. Contudo, uma série de demanda é uma série temporal que possui não linearidades e componentes periódicos aleatórios como o clima, perfil dos usuários, eventos públicos, economia, medições erradas, e, conseqüentemente, um modelo de previsão linear pode não ser apropriado. Este trabalho utiliza diferentes abordagens para formar comitês de *wavenets* para a previsão de séries temporais de demanda de energia elétrica de curto prazo, os desempenhos são comparados com uma rede neural artificial perceptron multicamadas com função de ativação sigmoide na camada oculta, uma *wavenet* simples, com a média da última semana e com o modelo inocente. As séries de demanda adotadas, isto é, duas séries de demanda anuais reais com medições horárias, passam por um estágio de pré-processamento para remoção da tendência e normalização, e também para transformação dos valores da série em conjuntos de entrada e saída para o treinamento supervisionado. Emprega-se a estratégia de previsão um passo à frente e a avaliação das previsões é realizada pelo coeficiente de correlação múltipla R^2 e também pela análise de correlação entre os resíduos. Para criação dos comitês utiliza-se a reamostragem com reposição, a validação cruzada e a dizimação de entradas, seleção construtiva, combinação pela média simples, moda, mediana e generalização empilhada. Os resultados dos testes de não linearidade demonstram que as duas séries consideradas são não lineares, e também constata-se a diminuição da assimetria dos dados após sua transformação. Do processo de seleção de variáveis obtém-se os atrasos máximos para cada série, valores passados que são utilizados como entradas, e percebe-se que são diferentes para cada série. O atraso máximo a ser utilizado como entrada tem influência na quantidade de amostras do conjunto de dados de entrada e saída. Uma característica dos resultados que se reflete em ambas as séries é o aumento do erro à medida que o horizonte de previsão aumenta. Os comitês de *wavenets* superam os demais modelos comparados, e, além do desempenho ser diferente para cada problema, o melhor método de aprendizado de comitê a ser utilizado também varia, bem como o horizonte de previsão máximo no qual os valores previstos se ajustam aos valores reais das séries. A qualidade das previsões é avaliada com testes de correlação dos resíduos.

Palavras-chave: *Wavenet*. Previsão de demanda de energia elétrica. Comitês. Redes neurais artificiais.

ABSTRACT

Electricity is part of a market which involves generation, transmission, distribution and consumption agents that aim their profit maximization and expenses minimization. To achieve that, they need a planning based on an accurate load forecast, since a pessimistic scenario may lead to more generators dispatch than needed, excessive reservoir and high operating costs, and, on the other hand, an optimistic scenario may place the electrical system at risk or requiring demand electricity purchasing on the free market for a very high price. Hence, load forecasting has been employed in areas such as optimal dispatch, maintenance planning, hydric reservoir planning, consumption pattern understanding, expansion planning, price forecasting and tax adjustments. However, a load series is a time series with nonlinearities and random periodic components as the weather, users profile, public events, economy and bad measures, therefore a pure linear model may not be appropriated. This work uses different approaches to create wavenet ensembles for short term load forecasting, the performances are compared with a multilayer perceptron with sigmoid activation function in the hidden layer, with a single wavenet, with the last week mean and also with the naive model. The load series adopted, that is, two annual hourly load series with actual measurements, are passed through a data pre-processing stage for trend removal and normalization, and also for conversion from the time series to a inputs and output set for supervised training. It is applied the one step ahead forecast strategy and the forecasting evaluation is accomplished by the multiple correlation coefficient, R^2 , and also by the residuals correlation analysis. For the ensemble creation are used the bootstrapping, cross-validation like, inputs decimation, constructive selection, simple average, median, mode and stacked generalization methods. The nonlinearity tests results demonstrate that both time series are nonlinear, and the asymmetry reduction after data transformation is verified. From the features selection process the maximum lags for each series are identified, lagged values to be used as inputs and it is noticed that they are different for each series. The maximum lag also influences the amount of samples in the dataset of inputs and outputs. A common characteristic of both series is that the error increase along with the prediction horizon. Results point out that the wavenets ensembles overcome the other compared models after tests with two actual annual hourly load series. Moreover, beyond the performance to be different for each problem, the best ensemble learning method also varies, as well as the maximum forecasting horizon for which the forecasted values fit the series actual values. The quality of the forecasts is analyzed through residuals correlation tests.

Key-words: Wavenet. Load forecasting. Ensembles. Artificial neural network.

LISTA DE FIGURAS

FIGURA 1 – DIAGRAMA DO PROCESSO DE PREVISÃO DE DEMANDA.....	16
FIGURA 2 – PLANO TEMPO-FREQUÊNCIA (T- Ω).....	23
FIGURA 3 – (A) FUNÇÃO CHAPÉU MEXICANO COM SUPORTE COMPACTO NO DOMÍNIO DO TEMPO E (B) NO DOMÍNIO DA FREQUÊNCIA.....	24
FIGURA 4 – <i>WAVELETS</i> NO PLANO TEMPO-FREQUÊNCIA.....	25
FIGURA 5 - ESTRUTURA DA <i>WAVENET</i>	26
FIGURA 6 – ETAPAS DA CONSTRUÇÃO DE UM <i>ENSEMBLE</i>	28
FIGURA 7 – PROPOSTA DE PESQUISA.....	47
FIGURA 8 - PREPARAÇÃO DOS DADOS DE TREINAMENTO PARA O <i>BAGGING</i>	53
FIGURA 9 - GERAÇÃO DE COMPONENTES VIA <i>BAGGING</i>	58
FIGURA 10 – FORMAÇÃO DOS CONJUNTOS DE TREINAMENTO E TESTE.....	59
FIGURA 11 - MÉTODO DA DIZIMAÇÃO DE ENTRADAS.....	60
FIGURA 12 – ESTIMAÇÃO DA DENSIDADE KERNEL.....	63
FIGURA 13 – GENERALIZAÇÃO EMPILHADA.....	65
FIGURA 14 - SÉRIE DE DEMANDA HORÁRIA GEFCOM NO ANO DE 2005.....	66
FIGURA 15 - SÉRIE DE DEMANDA HORÁRIA DA ITÁLIA NO ANO DE 2005.....	67
FIGURA 16 - TESTE DE NÃO LINEARIDADE PARA A SÉRIE GEFCOM.....	68
FIGURA 17 - TESTE DE NÃO LINEARIDADE PARA A SÉRIE ITÁLIA.....	68
FIGURA 18 - HISTOGRAMA DA SÉRIE GEFCOM (A) ANTES E (B) APÓS A TRANSFORMAÇÃO.....	69
FIGURA 19 - HISTOGRAMA DA SÉRIE ITÁLIA (A) ANTES E (B) APÓS A TRANSFORMAÇÃO.....	69
FIGURA 20 - SÉRIES DE DEMANDA HORÁRIA TRANSFORMADAS E NORMALIZADAS (A) GEFCOM E (B) ITÁLIA.....	70
FIGURA 21 - SERR PARA A SÉRIE GEFCOM.....	71
FIGURA 22 - SERR PARA A SÉRIE ITÁLIA.....	71
FIGURA 23 - ÍNDICE R ² MÉDIO DOS COMITÊS PARA A SÉRIE GEFCOM.....	74
FIGURA 24 - ETAPA DE SELEÇÃO PARA A SÉRIE GEFCOM.....	75
FIGURA 25 - COMPARAÇÃO ENTRE OS MÉTODOS DE COMBINAÇÃO PARA A SÉRIE GEFCOM.....	76
FIGURA 26 - ÍNDICE R ² MÉDIO DOS COMITÊS PARA A SÉRIE ITÁLIA.....	77
FIGURA 27 - PREVISÃO PARA A SÉRIE ITÁLIA COM E SEM A ETAPA DE SELEÇÃO.....	78
FIGURA 28 - COMITÊS DE <i>WAVENETS</i> COM DIFERENTES MÉTODOS DE COMBINAÇÃO PARA PREVISÃO DA SÉRIE ITÁLIA.....	79
FIGURA 29 - COMPARAÇÃO DO COMITÊ DE <i>WAVENETS</i> COM OUTROS ALGORITMOS PARA PREVISÃO DA SÉRIE GEFCOM.....	80
FIGURA 30 - COMPARAÇÃO DO COMITÊ DE <i>WAVENETS</i> COM OUTROS ALGORITMOS PARA PREVISÃO DA SÉRIE ITÁLIA.....	80
FIGURA 31 - TESTES DE CORRELAÇÃO PARA OS RESÍDUOS DE PREVISÃO DA SÉRIE GEFCOM.....	81

FIGURA 32 - TESTES DE CORRELAÇÃO PARA OS RESÍDUOS DE PREVISÃO DA SÉRIE ITÁLIA.....	82
FIGURA 33 - AMOSTRA DE PREVISÃO PARA A SÉRIE GEFCOM	83
FIGURA 34 - AMOSTRA DE PREVISÃO PARA A SÉRIE ITÁLIA	83

LISTA DE TABELAS

TABELA 1 - ESTATÍSTICA DA TRANSFORMAÇÃO BOX-COX SOBRE AS SÉRIES	67
TABELA 2 - VARIÁVEIS SELECIONADAS PARA CADA SÉRIE	72

LISTA DE SIGLAS

AE	- <i>Absolut Error</i>
ANN	- <i>Artificial Neural Network</i>
BEM	- <i>Basic Ensemble Method</i>
BMA	- <i>Bayesian Model Averaging</i>
CLSP	- <i>Cluster-based Subspace Projection</i>
DDS	- <i>Data from Different Sensors</i>
ENTSO-E	- <i>European Network of Transmission System Operators</i>
GASEN	- <i>Genetic Algorithm based Selective Neural Network Ensemble</i>
GBRT	- <i>Gradient Boosting Regression Trees</i>
GEM	- <i>Generalized Ensemble Method</i>
LEA	- <i>Layered Ensemble Architecture</i>
MAE	- <i>Mean Absolute Error</i>
MAPE	- <i>Mean Absolute Percentage Error</i>
MASE	- <i>Mean Absolute Scaled Error</i>
MLP	- <i>Multi Layer Perceptron</i>
MRE	- <i>Mean Relative Error</i>
MSE	- <i>Mean Square Error</i>
MSFE	- <i>Mean Square Forecast Error</i>
NCL	- <i>Negative Correlation Learning</i>
NFL	- <i>No Free Lunch theorem</i>
NMAE	- <i>Normalized Mean Absolute Error</i>
NMSE	- <i>Normalized Mean Square Error</i>
PCA	- <i>Principal Components Analysis</i>
RF	- <i>Random Forests</i>

RMS	- <i>Root Mean Square</i>
RNCL	- <i>Regularized Negative Correlation Learning</i>
SARIMA	- <i>Seasonal Autoregressive Moving Average</i>
SARIMAX	- <i>Seasonal Autoregressive Moving Average with Exogenous inputs</i>
SITB	- <i>Systems Identification Toolbox</i>
SFS	- <i>Sequential Forward Selection</i>
SVM	- <i>Support Vector Machines</i>
UCI	- <i>University of California Irvine</i>
WFT	- <i>Windowed Fourier Transform</i>
WNN	- <i>Wavelet Neural Network</i>

SUMÁRIO

1	INTRODUÇÃO.....	15
1.1	PROBLEMATIZAÇÃO.....	15
1.2	OBJETIVOS.....	17
1.2.1	Hipótese.....	18
1.2.2	Contribuições da dissertação.....	18
1.2.3	Organização da dissertação	19
2	REVISÃO DE LITERATURA.....	20
2.1	PREVISÃO DE DEMANDA DE CURTO PRAZO.....	20
2.2	WAVENETS.....	21
2.3	MÁQUINAS DE COMITÊ.....	26
2.3.1	Geração de componentes.....	28
2.3.2	Seleção de componentes	36
2.3.3	Combinação de componentes	41
2.4	PRÉ-PROCESSAMENTO DOS DADOS	45
3	MATERIAIS E MÉTODOS	46
3.1	PROCEDIMENTOS METODOLÓGICOS	46
3.2	PRÉ-PROCESSAMENTO DOS DADOS.....	48
3.2.1	GEFCom 2012.....	48
3.2.2	Itália	49
3.2.3	Transformação Box-Cox.....	49
3.2.4	Teste de não linearidade	49
3.2.5	Cálculo da janela de predição.....	50
3.2.6	Conversão da série temporal em dados de entrada e saída	52
3.3	SELEÇÃO DE VARIÁVEIS	53

3.4 ESTIMAÇÃO DAS WAVENETS	54
3.5 MÉTODO DE PREVISÃO	55
3.6 VALIDAÇÃO DOS MODELOS.....	56
3.7 MÉTODOS DE GERAÇÃO DE COMPONENTES PARA O COMITÊ.....	57
3.7.1 Reamostragem com reposição (<i>bootstrapping</i>)	57
3.7.2 Validação cruzada (<i>cross-validation</i>)	58
3.7.3 Dizimação de entradas (<i>inputs decimation</i>)	59
3.8 MÉTODO PARA SELEÇÃO DE COMPONENTES PARA O COMITÊ	60
3.9 MÉTODOS DE COMBINAÇÃO DE COMPONENTES DO COMITÊ	61
3.9.1 Média simples (<i>Basic Ensemble Method - BEM</i>)	61
3.9.2 Moda.....	62
3.9.3 Mediana.....	63
3.9.4 Generalização empilhada (<i>stacked generalization</i>)	63
3.10 DIFERENÇA ENTRE GENERALIZAÇÃO EMPILHADA, <i>BAGGING</i> E <i>ENSEMBLE</i>	64
4 RESULTADOS E DISCUSSÃO	66
4.1 PRÉ-PROCESSAMENTO DOS DADOS	66
4.2 COMITÊS DE WAVENETS.....	72
4.2.1 Previsão para a série GEFCom	73
4.2.2 Previsão para a série Itália	76
4.3 COMPARAÇÃO COM OUTROS ALGORITMOS.....	78
4.4 TESTES DE CORRELAÇÃO	81
5 CONCLUSÕES.....	84
REFERÊNCIAS.....	87

1 INTRODUÇÃO

1.1 PROBLEMATIZAÇÃO

A energia elétrica faz parte de um mercado que envolve agentes de geração, transmissão, distribuição e consumo, os quais por sua vez desejam maximizar seus lucros e minimizar as despesas. Para isso, é necessário um bom planejamento a partir de uma previsão de demanda precisa, já que uma previsão pessimista pode levar ao emprego de mais geradores que o necessário, reserva excessiva de matéria prima e aumento do custo de operação, por outro lado uma previsão otimista pode colocar o sistema elétrico em risco ou exigir a compra de energia no mercado livre a um preço alto. Tais problemas são considerados por Rana e Koprinska (2016) bem como por Raza e Khosravi (2015).

Por isso a previsão de demanda tem sido empregada no despacho de geradores, planejamento da manutenção, planejamento da reserva hídrica, compreensão do padrão de consumo, planejamento da expansão, previsão de preços e ajuste de tarifas, conforme exemplos encontrados em Rana e Koprinska (2016), Hong e Fan (2015), Raza e Khosravi (2015), Dudek (2011) e Cheng *et al.* (2012). O estado da arte de redes elétrica inteligentes, ou *smart grids*, elaborado por El-Hawary (2014) inclui a tarefa de previsão de demanda de energia elétrica como necessária para acelerar seu desenvolvimento, principalmente em tarefas como o gerenciamento em tempo real de consumidores residenciais, alocação ótima de recursos e tomada de decisão. A FIGURA 1 ilustra de forma geral estas aplicações.

Uma série de demanda é uma série temporal que possui não-linearidades e componentes periódicos aleatórios como o tempo, perfil dos usuários, eventos especiais, economia e medições erradas e por isso um modelo linear não consegue reproduzir toda a informação contida na série, o que nos induz a buscar um modelo que seja capaz desta tarefa.

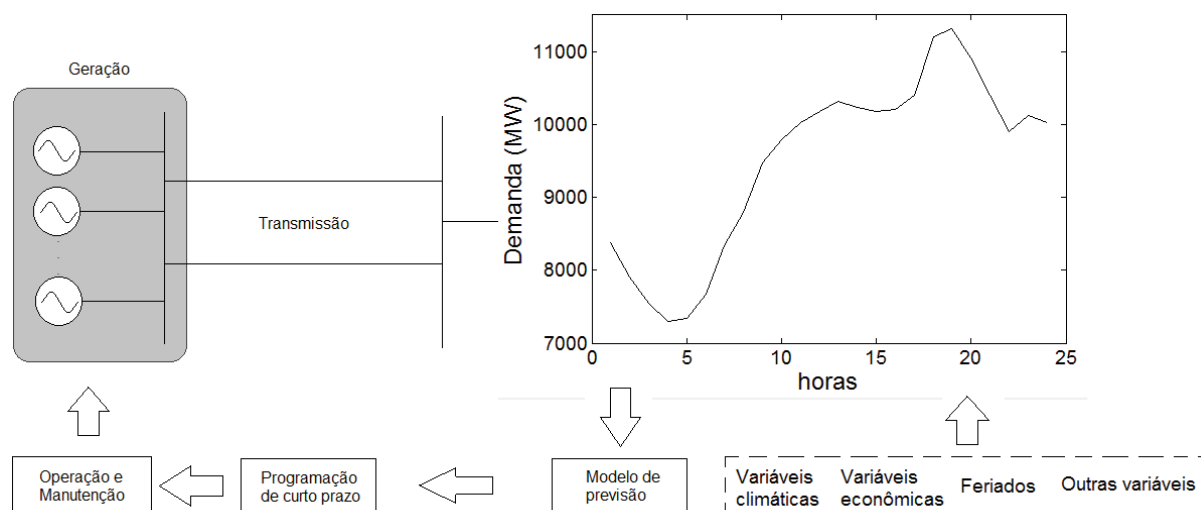


FIGURA 1 – DIAGRAMA DO PROCESSO DE PREVISÃO DE DEMANDA

FONTE: O autor (2016)

Contudo, Wolpert e MacReady (1997) provaram através do teorema “*No free lunch*” (NFL) que não há um algoritmo de aprendizado melhor do que qualquer outro para todos os problemas em um determinado espaço, sugerindo que a comparação entre diferentes algoritmos é desnecessária. Contudo, Zhou (2012) contra argumenta que em problemas reais não se faz necessário encontrar o modelo que seja melhor em todos os problemas do espaço considerado, sendo suficiente ser melhor para o problema específico a ser resolvido, justificando a busca por tal modelo.

Os comitês combinam vários modelos para resolver um problema específico e, portanto espera-se que seja melhor do que os modelos individuais em uma área maior no domínio do problema, mas ainda assim não se espera que seja melhor em todo o espaço (Kuncheva, 2014), já que a precisão em domínios do mundo real pode ser aumentada à custa da perda de precisão em domínios que nunca ocorrerão na prática (Rokach, 2010). Lima (2004) e Cavalcanti *et al.* (2016) também justificam o emprego de comitês para tentar vencer o teorema NFL. Apesar do pessimismo intrínseco ao teorema NFL, é possível aprender que o esforço para tentar encontrar um algoritmo universal que resolva todos os problemas possíveis é em vão.

Encontram-se na literatura redes neurais artificiais utilizadas como modelos de previsão de séries temporais que possuem não linearidades em maior ou menor

grau dependendo do problema tratado, contudo o problema ainda não está superado e algumas redes neurais artificiais falham nas suas previsões devido a alguns fatores, como a convergência para mínimos locais, pelo sobreajuste do ruído nos dados (*overfitting*) ou pelo conjunto de dados de treinamento limitado, portanto a combinação de modelos imperfeitos em um comitê é uma maneira de minimizar estas deficiências individuais (Ahmad e Zhang, 2009).

Como exemplo de aplicação de comitês, De Felice e Yao (2011) utilizaram um comitê de redes neurais artificiais para prever a demanda de energia elétrica em oito edifícios comerciais e em todos foi constatado que o comitê obteve melhor previsão do que seu pior preditor individual, sendo que em um dos casos a previsão do comitê foi melhor do que a de seu melhor componente. Os comitês também foram aplicados na previsão de velocidade do vento por Graff *et al.* (2014) e foi constatado que seu desempenho foi superior a de qualquer previsão individual.

O problema de previsão pode ser assim resumido: dada uma série temporal de demanda X_1, X_2, \dots, X_t , em que X_i é o valor da demanda no instante i , registrada a um intervalo fixo até o instante t , a tarefa é prever X_{t+h} , onde h é o horizonte de previsão. Este trabalho investiga a utilização de comitês de uma classe específica de redes neurais artificiais, as *wavenets*, para a previsão de demanda de energia elétrica de curto prazo, ou seja, de uma até 168 horas à frente.

1.2 OBJETIVOS

O foco deste trabalho é investigar a previsão de sistemas com o objetivo de estudar situações reais complexas através de metaheurísticas para identificação de sistemas não lineares.

Para se atingir o objetivo geral são perseguidos os seguintes objetivos específicos:

- a) analisar algoritmos de aprendizado de comitês existentes na literatura;
- b) tratar os conjuntos de dados para adequá-los aos algoritmos de criação dos modelos de previsão;

- c) implementar os algoritmos de comitês de *wavenets* e outros existentes para obter os modelos preditores;
- d) obter as previsões e calcular índices de desempenho e validação dos modelos;
- e) comparar e discutir os resultados.

1.2.1 Hipótese

Tem-se como hipótese que a previsão de demanda de curto prazo é mais precisa ao se utilizar comitês de *wavenets* como modelo preditor ao invés de *wavenets* individuais, redes neurais artificiais individuais, a média ou o modelo inocente.

1.2.2 Contribuições da dissertação

Pode-se citar como primeira contribuição desta dissertação a revisão da literatura sobre o tema de previsão de demanda utilizando máquinas de comitê. Outra contribuição é a sistematização dos métodos selecionados para experimentação.

Do teorema NFL sabe-se que não há um modelo universal para resolver todos os problemas possíveis. Esta dissertação aplica comitês de *wavenets* em problemas de previsão de demanda de energia elétrica e os resultados indicam que tais comitês são melhores do que alguns outros modelos comparados no âmbito dos problemas experimentados.

Explora-se neste trabalho a potencialidade das *wavenets* de capturar as características não lineares de sinais ricos em frequência, como é o caso de séries temporais de demanda de energia elétrica, que possuem frequências anuais, mensais, semanais e diárias.

Além do mais, ao se transferir o foco da identificação de um modelo de previsão da otimização de um modelo específico para a interação entre vários

modelos, ou seja, comitês de máquinas, que não necessitam de modelos tão perfeitos, mas que sejam divergentes e com boa aproximação, torna-se possível obter previsões melhores sem se preocupar com questões como a melhor estrutura, parâmetros internos, inicialização, dentre outros. As características das *wavenets* que corroboram com os comitês de máquinas, que buscam integrar muitos modelos, é que sua inicialização não é aleatória como ocorre nas demais redes neurais artificiais, isto implica em menos iterações para o aprendizado, e, além disso, preservam a propriedade de aproximação universal das redes neurais artificiais (Alexandridis e Zaprani, 2013; Thuillard, 2002). Outra característica das *wavenets* que contribui para os comitês é que elas são consideradas instáveis, ou seja, seu comportamento muda de acordo com os exemplos apresentados durante seu aprendizado, e no contexto de comitês de máquinas é considerada uma vantagem já que permite introduzir divergência entre os modelos.

1.2.3 Organização da dissertação

No capítulo 1 desta dissertação foram introduzidos a problematização da pesquisa, os objetivos gerais, específicos e a hipótese.

No capítulo 2 é apresentada uma revisão de literatura dos principais tópicos abordados nesta dissertação, os quais são a previsão de séries temporais, em especial a previsão de demanda de curto prazo, *wavenets*, máquinas de comitê e, por fim, o pré-processamento de dados.

Os materiais e métodos utilizados estão contemplados no capítulo 3 que inclui a exposição dos conjuntos de dados e seu pré-processamento, seleção de variáveis, estimação das *wavenets*, métodos de previsão, métodos de validação dos modelos preditores e os métodos para aprendizado de comitê.

Os experimentos e seus respectivos resultados são demonstrados e discutidos no capítulo 4, e as conclusões desta dissertação estão no capítulo 5.

2 REVISÃO DE LITERATURA

Neste capítulo são apresentadas as principais referências encontradas na literatura nas áreas de previsão de demanda, *wavenets*, máquinas de comitê e de pré-processamento de dados para modelagem. As referências são apresentadas preferencialmente na ordem cronológica, exceto quando há trabalhos similares publicados em diferentes anos.

2.1 PREVISÃO DE DEMANDA DE CURTO PRAZO

A previsão de curto prazo é considerada no período de uma hora até 168 horas à frente, ou seja, uma semana. Esta seção apresenta alguns trabalhos encontrados na literatura que abordam o problema em questão. Em Gao e Tsoukalas (2001) é aplicada uma *wavenet* para a previsão de demanda em dois estágios, primeiramente classificando o padrão da carga como residencial, comercial, industrial grande ou pequeno (esta etapa é denominada pelos autores como identificação da demanda), e posteriormente é realizada a previsão da demanda como um problema de regressão.

Em Kermanshahi e Iwamiya (2002) considera-se que os fatores climáticos como a temperatura afetam em maior grau as previsões de curto prazo, enquanto que as previsões de longo prazo são afetadas em maior grau por fatores econômicos, e aplicaram redes neurais artificiais para a previsão de picos de demanda no Japão até o ano de 2020.

As principais questões e evidências sobre os fundamentos da previsão de demanda de energia elétrica de curto prazo são abordadas por Aguirre *et al.* (2008), onde o problema é separado em dois: mapeamento de padrões e previsão dinâmica. Os autores concluem que a não linearidade da série temporal é dependente do problema e por vezes pode ser obtido um modelo através do mapeamento de um padrão seguido da identificação de um modelo linear e em outros casos a não

linearidade é acentuada e as redes neurais artificiais são uma alternativa, principalmente na modelagem caixa-preta.

De Felice e Yao (2011) aplicaram comitês de redes neurais artificiais para a previsão de demanda em oito edifícios comerciais em Roma, Itália. As redes neurais artificiais também são utilizadas na previsão dos custos da energia elétrica no mercado livre conforme pode ser visto nos trabalhos de Andalib e Atry (2009), Areekul *et al.* (2011) e Assareh *et al.* (2012). Uma revisão dos modelos para previsão de demanda foi elaborada por Suganthi e Samuel (2012).

A previsão do preço de energia de curto prazo foi abordada por Hu e Taylor (2014) ao utilizar modelos *Fuzzy-C-Means* e máquinas de vetores suporte (SVM) para esta tarefa utilizando dados do Reino Unido de 2007. Recentemente Nadtoke e Balasim (2015) aplicaram máquinas de vetores suporte (SVM) e enxame de partículas (PSO) para previsão do consumo de eletricidade levando em consideração fatores meteorológicos como a temperatura do ar e a iluminação natural.

Já Papadopoulos e Karakatsanis (2015) compararam o desempenho de comitês derivados dos métodos *Random Forests* (RF) e *Gradient Boosting Regression Trees* (GBRT) com os modelos convencionais auto regressivo com média móvel sazonal (SARIMA) e auto regressivo com média móvel sazonal com variáveis exógenas (SARIMAX).

O QUADRO 1 apresenta as principais publicações, algoritmos e abordagens relacionadas ao problema de previsão de demanda de energia elétrica.

2.2 WAVENETS

Existem aplicações em que a relação entre o instante de tempo e a frequência do sinal neste instante de tempo é necessária, como por exemplo, as notas musicais, porém a transformada de Fourier de um sinal limitado no tempo não é limitada no domínio da frequência, e vice-versa, ou seja, apenas com a transformada de Fourier não é possível obter a relação direta no plano tempo-frequência.

Autores	Algoritmo	Abordagem / Aplicação
Hu e Taylor (2014)	<i>Fuzzy C means</i> Máquinas de Vetores Suporte	Preço da energia de curto prazo
Suganthi e Samuel (2012)	Revisão de algoritmos	Previsão de demanda
Nadtoka e Balasim (2015)	SVM e PSO	
Papadopoulos e Karakatsanis (2015)	SARIMA, SARIMAX, RF, GBRT	
De Felice e Yao (2011)	Comitês de redes neurais artificiais	Regressão
Andalib e Atry (2009) Areekul <i>et al.</i> (2011) Assareh <i>et al.</i> (2012)	Redes neurais artificiais	Previsão de preço da energia elétrica
Aguirre <i>et al.</i> (2008)	NARMA e MLP	Mapeamento de padrões + Previsão dinâmica
Kermanshahi e Iwamiya (2002)	Redes neurais artificiais	Picos de demanda anuais
Gao e Tsoukalas (2001)	<i>wavenet</i>	Classificação + Regressão

QUADRO 1 - PUBLICAÇÕES, ALGORITMOS E ABORDAGENS RELACIONADAS AO PROBLEMA DE PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA DE CURTO PRAZO

FONTE: O autor (2016)

De acordo com Daubechies (1993), esta relação direta entre tempo e frequência pode ser obtida pela transformada de Fourier janelada (*windowed Fourier transform* - WFT), conforme ilustrado na FIGURA 2, de modo que

$$(T^{win}f)(\omega, t) = WFT = \int_{-\infty}^{+\infty} f(t) g(t-s) e^{-i\omega t} dt \quad (2.1)$$

onde a função $g(t-s) e^{-i\omega t}$ é denominada átomo, possui energia finita, suporte compacto, frequência central ω , e simetria com relação ao instante s , $f(t)$ é a função que está sendo transformada em função do tempo t .

Ao se analisar a FIGURA 2 é possível identificar duas WFT para funções átomos do tipo gaussianas, $g(t-u)e^{-i\xi t}$ e $g(t-v)e^{-i\gamma t}$, simétricas no domínio do tempo em relação a u e v , e frequência central em ξ e γ respectivamente. Devido ao princípio da incerteza de Heisenberg a relação biunívoca tempo-frequência é dada através da caixa de Heisenberg delimitada pelo desvio padrão das funções no domínio do tempo e no domínio da frequência (Chaplais, 2015).

Daubechies (1993) introduz a *wavelet* como sendo uma classe das funções de janelamento utilizadas na WFT, porém que possuem um suporte variável de acordo com a frequência do sinal, tal que

$$(T^{wav}f)(\omega, t) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \varphi\left(\frac{t-b}{a}\right) dt \quad (2.2)$$

onde $\varphi\left(\frac{t-b}{a}\right)$ é a *wavelet*-mãe, b é o parâmetro de translação e a é o parâmetro de escala ou dilatação. Uma escolha típica para a *wavelet*-mãe é a função chapéu mexicano (*mexican hat*), ilustrada na FIGURA 3, tal que

$$\varphi(t) = (1 - t^2)e^{-t^2/2} \quad (2.3)$$

A representação de duas *wavelets* no plano tempo-frequência é dada na FIGURA 4, na qual é possível observar que a translação no domínio do tempo funciona do mesmo modo que na WFT, porém o parâmetro de escala produz um alargamento (ou achatamento) na caixa de Heisenberg de acordo com a frequência, aumentando a resolução no domínio da frequência para altas frequências e aumentando a resolução no domínio do tempo para baixas frequências, ao contrário da WFT que mantém a resolução (escala) constante independentemente da frequência.

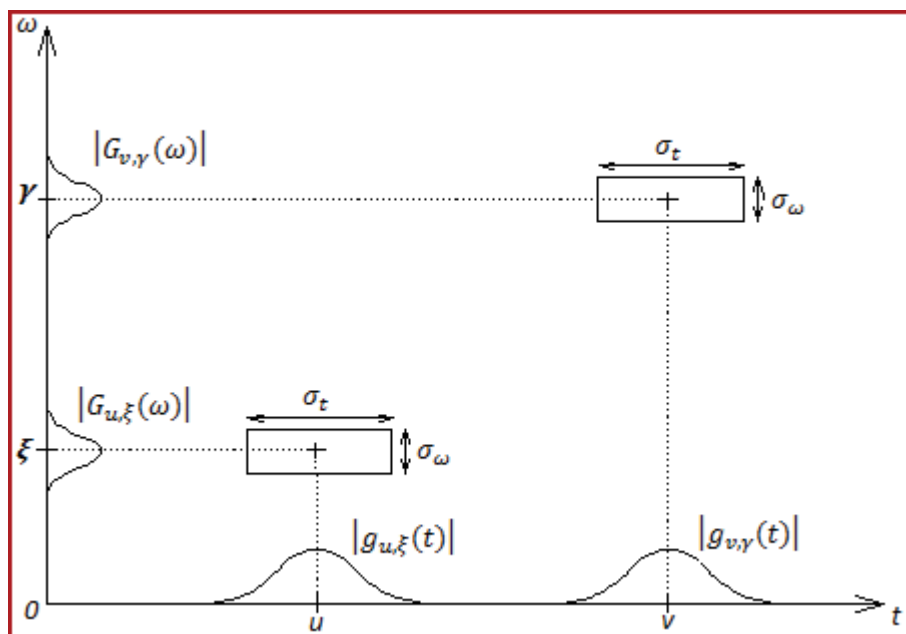


FIGURA 2 – PLANO TEMPO-FREQÜÊNCIA ($t - \omega$)

FONTE: Chaplais (2015)

Alexandridis e Zapanis (2013) definem brevemente a *wavenet* (*wavelet neural network*) como uma rede com uma camada escondida que possui *wavelets*

como função de ativação, a qual tem sua estrutura representada na FIGURA 5 e foram propostas inicialmente por Zhang e Benveniste (1992). A *wavenet* possui m entradas $(x_1 \dots x_m)$, λ neurônios na camada escondida denominados *wavelons*, m conexões diretas entre as entradas e o neurônio de saída para modelar a parte linear do sistema com os pesos $\{w_1^{[0]} \dots w_m^{[0]}\}$, as conexões entre as entradas e os *wavelons* correspondem aos pesos de translação ξ e dilatação ζ das wavelets dado pelo conjunto $w_{ij}^{[1]} = \{w_{(\xi)ij}^{[1]}, w_{(\zeta)ij}^{[1]}\}$, em que $\{i, j \in Z \mid 1 \leq i \leq m \text{ e } 1 \leq j \leq \lambda\}$.

A saída do *wavelon* j é dada por

$$\Psi_j(x) = \prod_{i=1}^m \psi\left(\frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}}\right) \quad (2.4)$$

onde i e j representam o número da entrada e do *wavelon* respectivamente dado que $\{i, j \in Z \mid 1 \leq i \leq m \text{ e } 1 \leq j \leq \lambda\}$. As conexões entre os *wavelons* e o neurônio de saída são dadas pelos pesos $\{w_1^{[2]} \dots w_\lambda^{[2]}\}$ e, além dessas conexões, há uma conexão extra para correção da polarização (*bias*) dada pelo peso $w_{\lambda+1}^{[2]}$ multiplicado pela unidade.

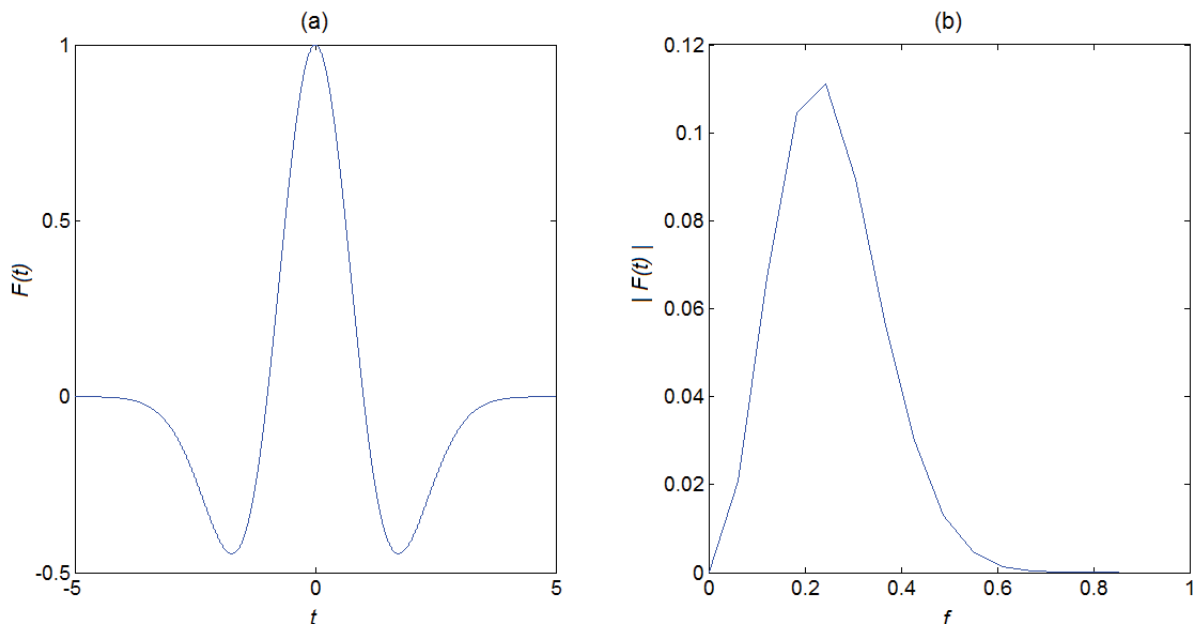


FIGURA 3 – (A) FUNÇÃO CHAPÉU MEXICANO COM SUPORTE COMPACTO NO DOMÍNIO DO TEMPO E (B) NO DOMÍNIO DA FREQUÊNCIA

FONTE: O autor (2016)

A saída da *wavenet* é dada por

$$\hat{y}(x) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \prod_{i=1}^m \psi \left(\frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}} \right) + \sum_{i=1}^m w_i^{[0]} \cdot x_i \quad (2.5)$$

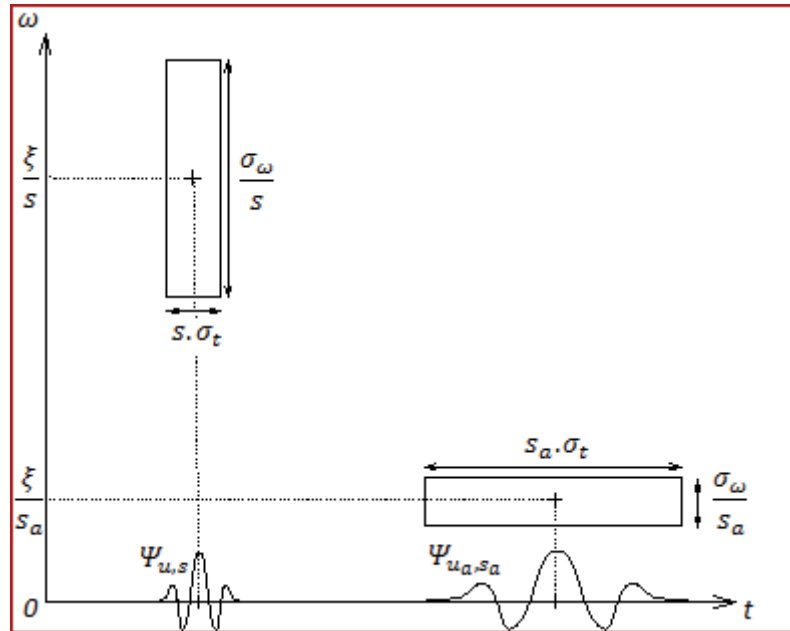


FIGURA 4 – WAVELETS NO PLANO TEMPO-FREQUÊNCIA

FONTE: Chaplais (2015)

De acordo com Kermanshahi e Iwamiya (2002), se o número de *wavelons* for menor que o necessário a correlação entre entrada e saída não é bem compreendida pela rede, por outro lado se a quantidade de neurônios for maior do que o necessário até os ruídos são modelados e entendidos como correlação entre entrada saída, diminuindo sua capacidade de generalização. Conclui-se que a determinação da estrutura da *wavenet* é dependente do problema em estudo e deve ser determinada caso a caso.

As *wavenets* têm sido aplicadas em diversas áreas do conhecimento, tais como, engenharia financeira, teoria do caos (Alexandridis e Zaprani, 2014), aproximação de funções em Awad (2012), previsão de falência de bancos em Chauhan *et al.* (2009), no ambiente automotivo por Klein (2011) e na previsão do consumo de condicionadores de ar em Liao (2014).

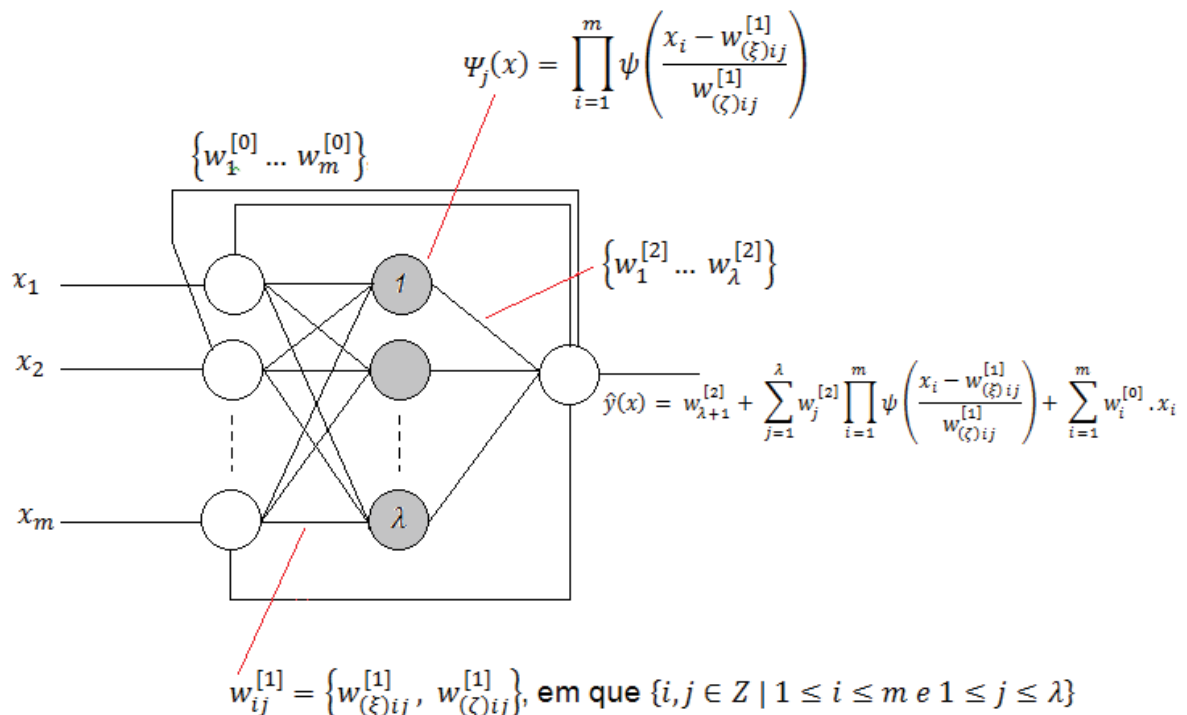


FIGURA 5 - ESTRUTURA DA WAVENET

FONTE: Adaptado de Alexandridis e Zaprani (2013)

2.3 MÁQUINAS DE COMITÊ

Um dos primeiros trabalhos a utilizar comitês de redes neurais artificiais foi o de Hansen e Salamon (1990), desde então surgiram diversos algoritmos para o aprendizado de tais comitês. Nesta seção são apresentados brevemente os algoritmos de aprendizado de comitês encontrados na literatura, em ordem cronológica, exceto quando é conveniente agrupar algoritmos similares de épocas diferentes, e ao final são organizados nas seguintes categorias de acordo com a classificação utilizada por Moreira *et al.* (2012):

- a) geração;
 - i. manipulação de dados;
 - A. reamostragem;
 - B. manipulação do espaço de características;

- C. manipulação das saídas;
- ii. manipulação da estimação dos modelos;
 - A. parâmetros;
 - B. algoritmo de indução;
 - abordagem sequencial;
 - abordagem paralela;
- b) seleção ou poda;
 - i. exponencial;
 - ii. aleatória;
 - iii. sequencial;
 - A. seleção para frente;
 - B. seleção para trás;
 - C. seleção mista para frente e para trás;
 - D. ordenação;
 - E. aglomeração (*clustering*);
- c) combinação ou integração;
 - i. pesos constantes;
 - ii. pesos não constantes;
 - A. áreas de especialização;
 - B. abordagem dinâmica.

Para Lima (2004), a construção de um comitê de máquinas se resume em três passos básicos, que são a geração, seleção (opcional, porém proporciona melhores resultados) e combinação de componentes, os quais demandam três conjuntos de dados com mesma distribuição amostral, sendo um conjunto para geração de componentes, outro para seleção e combinação, e o último para teste de desempenho do comitê.

Um *ensemble* efetivo é aquele que tem um desempenho melhor do que o de seu melhor componente tomado isoladamente, e esta efetividade é maior quanto menor for o grau de correlação entre seus erros individuais. Existem três abordagens que podem ser utilizadas (Lima, 2004):

- escolher um método de combinação que considere a dependência entre os componentes;
- gerar os componentes descorrelacionados;
- selecionar componentes descorrelacionados.

A FIGURA 6 ilustra as etapas da construção de um comitê: geração, seleção e combinação, na qual x_1, x_2 são entradas, $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m$ são as previsões dos modelos componentes do comitê e \bar{y} é o resultado da combinação destas previsões.

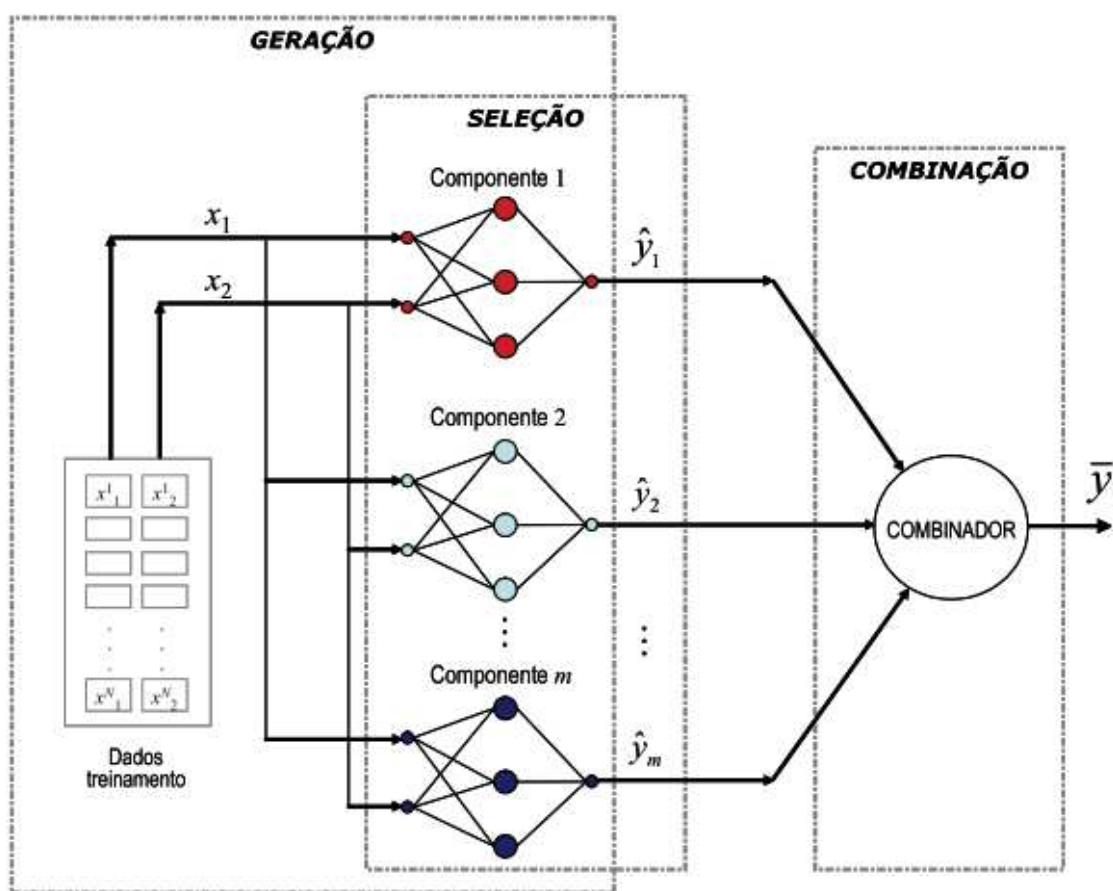


FIGURA 6 – ETAPAS DA CONSTRUÇÃO DE UM *ENSEMBLE*

FONTE: Adaptado de Villanueva (2006)

2.3.1 Geração de componentes

De acordo com Lima (2004) o objetivo da fase de geração é encontrar componentes com bom desempenho individual e que generalizem de forma

descorrelacionada dos outros, o que pode ser obtido de forma empírica ou através da maximização da divergência. Na forma empírica são alterados os parâmetros de treinamento da rede neural, como a condição inicial, os dados de treinamento, a topologia da rede e o algoritmo de treinamento. Já para o incremento da divergência podem ser realizados o pré-processamento dos parâmetros e estrutura da rede neural artificial, o pré-processamento dos dados de treinamento ou a variação do método de busca no espaço de hipóteses (Barbosa, 2009).

Os métodos de geração de componentes para comitês encontrados na literatura são apresentados no QUADRO 2.

2.3.1.1 Manipulação de dados: reamostragem

O *bagging* foi proposto originalmente no trabalho de Breiman (1996) e se resume a realização de várias amostragens aleatórias com reposição (*bootstrap*) sobre um conjunto de dados de treinamento, geração de um preditor para cada conjunto e sua posterior combinação através de média simples, formando o termo *bootstrap aggregating*, ou *bagging*. A diferenciação dos conjuntos de treinamento tende a produzir preditores que generalizam de forma diferente. O *bagging* também foi utilizado por Fan *et al.* (2009) na previsão de demanda de curto prazo de uma zona de consumo norte-americana utilizando dados meteorológicos comerciais e por Khwaja *et al.* (2015) na previsão de demanda de energia elétrica de curto prazo com redes neurais artificiais.

O *boosting* (Schapire, 1990) é uma reamostragem adaptativa em que as amostras que mais contribuem para o erro de treinamento dos componentes têm maior probabilidade de participar do treinamento do próximo componente a ser gerado e é aplicado geralmente em problemas de classificação. O Adaboost.R, proposto por Freund e Schapire (1996), é uma modificação do boosting para problemas de regressão.

ALGORITMO	PUBLICAÇÃO	CATEGORIA
<i>Bagging</i>	Breiman (1996)	Manipulação de dados: reamostragem
<i>AdaBoost.R</i>	Freund e Schapire (1996)	
<i>k-nearest neighbors</i>	Drezga e Rahman (1999)	
<i>Gradient Boosting</i>	Friedman (2001)	
Validação cruzada	Wichard e Ogorzalek (2004)	
CLSP	Hadavandi <i>et al.</i> (2015)	Manipulação dos dados: reamostragem e espaço de características
PSEL	Yu <i>et al.</i> (2016)	
Transformação do subespaço: comitês com ruído	Raviv e Intrator (1996)	Manipulação de dados: espaço de características
<i>Random Subspace</i>	Ho (1998)	
Dizimação de entradas	Tumer e Ghosh (1996)	
Seleção do subespaço de características	Opitz (1999)	
	Zenobi e Cunningham (2001)	
	Domeniconi e Yan (2004)	
<i>Random Forest</i>	Breiman (2001)	
Transformação do subespaço: ruído nas entradas	Frank e Pfahringer (2006)	
<i>Rotation Forest</i>	Rodriguez <i>et al.</i> (2006)	
Previsões irmãs	Nowotarski <i>et al.</i> (2016)	
Adição de ruído na saída	Breiman (2000)	Manipulação de dados de saída
<i>Bagging</i> iterativo	Breiman (2001)	
Diferente número de camadas ou de neurônios na camada escondida	Perrone e Cooper (1993)	Manipulação da estimação do modelo: parâmetros
Parâmetros iniciais aleatórios	Rosen (1996)	
Poda parcial de pesos de redes neurais artificiais	Chen <i>et al.</i> (2016)	
Redes neurais artificiais com penalidade por correlação	Rosen (1996)	Manipulação da estimação do modelo: algoritmo de indução, sequencial
CNNE	Islam <i>et al.</i> (2003)	
SECA	Granitto <i>et al.</i> (2005)	
CVM adaptado	Tsang <i>et al.</i> (2006)	
ADDEMUP	Opitz e Schavlik (1996b)	Manipulação da estimação do modelo: algoritmo de indução, paralelo
NCL	Liu e Yao (1999)	
EENCL	Liu <i>et al.</i> (2000)	
RTQRT-NCL	McKay e Abbass (2001)	
RNCL	Chen e Yao (2009)	
SNCL	Hadavandi <i>et al.</i> (2015)	

QUADRO 2 - ALGORITMOS DE GERAÇÃO DE COMPONENTES PARA COMITÊS

FONTE: O autor (2016)

Em Drezga e Rahman (1999) foi proposto um novo método de seleção da quantidade de neurônios escondidos de redes neurais artificiais e o treinamento em conjuntos de dados gerados através do algoritmo k-vizinhos mais próximos para

compor um *ensemble* aplicado na previsão de demanda de energia elétrica em duas plantas dos Estados Unidos.

Wichard e Orgozalek (2004) adaptaram o método da validação cruzada para gerar os componentes de *ensembles* utilizados na previsão de séries temporais por meio de preditores híbridos lineares e não lineares. Lee e Baldick (2014) empregaram um *ensemble* de redes neurais artificiais na previsão de geração eólica de sete parques eólicos.

2.3.1.2 Manipulação de dados: reamostragem e espaço de características

Hadavandi *et al.* (2015) propuseram um comitê de redes neurais artificiais denominado *Boosted-NNE*, o qual utiliza uma mistura do *boosting* com um novo método de projeção de subespaços por aglomeração denominado *cluster-based sub-space projection* (CLSP), o qual seleciona uma pequena porção do espaço de características gerado pelo *boosting*.

Recentemente foi proposto por Yu *et al.* (2016) um método que mescla a manipulação dos dados pela reamostragem e do espaço de características, alternando entre o *Random Subspace* e o *Adaboost*.

2.3.1.3 Manipulação de dados: espaço de características

A dizimação de entradas Tumer e Ghosh (1996) é um método que visa diminuir a correlação entre os preditores através da redução da dimensão do conjunto de observações na fase de treinamento, de forma controlada para que não haja um aumento significativo na polarização e variância do erro, induzindo pequenas diferenças no comportamento dos preditores individuais e consequentemente sua correlação.

Raviv e Intrator (1996) combinaram a reamostragem com reposição e a introdução de ruídos para gerar componentes em um comitê de redes neurais artificiais. Ho (1998) percebeu que os trabalhos que utilizavam as árvores de decisão

focavam sempre na otimização do tamanho das árvores e raramente na sua capacidade de generalização, por isso propôs o método *Random Subspace*, o qual toma variáveis aleatórias do espaço de entrada para construir diversos modelos que são combinados ao final.

Uma forma de gerar modelos com diversidade entre si é treinar cada um com um subespaço das variáveis de entrada. Ho (1998) utilizou uma seleção aleatória de um fator m (0 a 100%) definido como parâmetro do modelo para produzir K subconjuntos de dados a partir do espaço de entradas original, denominado *Random Subspace*.

Neste sentido Opitz (1999) aborda o problema de seleção de variáveis para o comitê ao assumir a hipótese que as variáveis ótimas para o comitê são diferentes das variáveis ótimas para cada modelo individual, ele utiliza algoritmos genéticos para a seleção de variáveis.

Zenobi e Cunningham (2001) também abordam a seleção ótima de variáveis para o comitê, no sentido de diminuir o erro de generalização e chega à conclusão de que a quantidade de variáveis para o comitê é menor do que quando se considera somente modelos individuais.

O algoritmo *Nearest Neighbors*, ou seja, vizinhos mais próximos, é robusto quanto à reamostragem dos dados, mas conforme estudado por Domeniconi e Yan (2004) o algoritmo é sensível à variação do espaço de características.

A transformação de um conjunto de amostras real em um artificial através da introdução de ruído no espaço de entrada é utilizada por Frank e Pfahringer (2006). O ruído tende a gerar conjunto de dados de treinamento diferentes e conseqüentemente modelos divergentes.

Geralmente existe um estágio de seleção de variáveis antes da estimação de um modelo, o qual seleciona apenas as variáveis que contribuem mais significativamente de acordo com algum critério. Este estágio visa reduzir a dimensão do espaço de entrada evitando sobreajuste ao reduzir a ordem do modelo bem como evitando esforço computacional desnecessário, contudo existe uma perda de informação associada às variáveis removidas. Para aproveitar todas estas informações foram desenvolvidos os métodos *Random Forest* (Breiman, 2001) e *Rotation Forest* por Rodriguez *et al.* (2006).

O *Random Forest* foi aplicado na previsão de demanda de energia elétrica de curto prazo por Dudek (2011) e Cheng *et al.* (2012). Já o *Rotation Forest* que foi criado para problemas de classificação foi adaptado para problemas de regressão por Lasota *et al.* (2012) e Pardo *et al.* (2013). A principal diferença entre os dois métodos é que o *Rotation Forest* rotaciona os dados de entrada utilizando o algoritmo PCA e cria os componentes sequencialmente enquanto que o *Random Forest* não rotaciona os dados.

O método *Random Forest* seleciona variáveis de entrada de forma aleatória para geração dos preditores base, já Nowotarski *et al.* (2016) propuseram um método que utiliza diferentes algoritmos de seleção que tentam obter as melhores variáveis para geração dos preditores base, contudo as variáveis selecionadas geralmente variam no resultado de um algoritmo de seleção para outro, introduzindo desta forma diversidade entre os modelos. As previsões dos modelos gerados a partir das variáveis selecionadas de cada algoritmo são denominadas de previsões irmãs.

2.3.1.4 Manipulação de dados de saída

Ao fazer a reamostragem dos dados com reposição como no *bagging* está se perturbando os dados de entrada e saída. Breiman (2000) realizou esta perturbação de forma independente nos dados de saída e obteve melhor desempenho do que o *bagging*.

Outro método que manipula as variáveis de saída é o *bagging* iterativo (Breiman, 2001), que cria modelos sequencialmente para compor o comitê. O primeiro modelo é criado através do *bagging*, para treinar o segundo modelo os dados de saída são os resíduos do primeiro modelo, e assim sucessivamente até que se atinja um critério de parada.

2.3.1.5 Manipulação dos parâmetros de estimação do modelo

Perrone e Cooper (1993) variaram a quantidade de neurônios na camada escondida de redes neurais artificiais para criar modelos divergentes para o comitê.

As redes neurais artificiais são sensíveis às condições iniciais, ou seja, caso os valores iniciais dos parâmetros de cada rede sejam gerados aleatoriamente é provável que após o treinamento as redes resultantes sejam divergentes. Com base nesta ideia Rosen (1996) criou um comitê de redes neurais artificiais descorrelacionadas.

Recentemente Chen *et al.* (2016) propuseram um novo método de geração de componentes para comitês que se inicia com a geração de redes neurais artificiais como modelos preditores através de um algoritmo genético que maximiza a divergência entre elas e, logo após, aplicam-se podas parciais em alguns pesos de cada preditor afim de se introduzir mais divergências.

2.3.1.6 Manipulação do algoritmo de indução do modelo: sequencial

Em Rosen (1996) as redes neurais artificiais em um comitê são treinadas não somente para obter a resposta desejada, mas também diminuir a correlação com as respostas das outras redes concorrentes.

Em Islam *et al.* (2003) é apresentado um algoritmo que determina simultaneamente a quantidade de redes neurais artificiais no *ensemble*, a quantidade de neurônios em cada uma, a quantidade de épocas de treinamento de cada rede neural e avalia o desempenho do *ensemble* através do seu próprio erro ao invés da decomposição polarização, variância e diversidade.

Outra abordagem sequencial é o algoritmo SECA proposto por Granitto *et al.* (2005). O algoritmo utiliza a reamostragem com reposição para gerar cada modelo, de forma que os modelos são gerados e adicionados ao comitê até que não haja mais melhoria no desempenho do comitê.

Para tratar de uma quantidade massiva de dados Tsang *et al.* (2006) propuseram um comitê de CVM, *Core Vector Machines*, com uma versão diversificada da original e constatou que o desempenho superou o comitê com CVM originais além de ser mais veloz.

2.3.1.7 Manipulação do algoritmo de indução do modelo: paralela

Em Opitz e Shavlik (1996a) é apresentado o algoritmo genético ADDEMUP que cria uma população inicial de redes e utiliza operadores genéticos para aumentar sua diversidade nas próximas gerações. O algoritmo foi testado em um problema de DNA em Opitz e Shavlik (1996b).

Liu e Yao (1999) propuseram o aprendizado simultâneo das redes neurais artificiais em um *ensemble* através da correlação negativa (NCL – *Negative Correlation Learning*), ou seja, todas as redes neurais artificiais são treinadas e a correlação entre elas em cada época é medida e seus pesos são ajustados para minimizar sua correlação, que é introduzida como uma penalidade na função custo. Barbosa (2009) argumenta que apesar do sucesso do NCL existem diversos trabalhos que observaram pouca diversidade nos *ensembles* criados por métodos de penalidade, argumentando que o mesmo afasta os componentes da média mas não entre si. Para tentar corrigir este problema McKay e Abbass (2001) utilizam o NCL com uma função de penalização alternativa.

Em Liu *et al.* (2000) apresenta-se um método para a construção do *ensemble* que utiliza o NCL em conjunto com um algoritmo de programação evolutiva que divide os indivíduos em espécies, sendo que cada espécie representa redes neurais artificiais com quantidades de neurônios diferentes. O algoritmo é nomeado EENCL.

Chen e Yao (2009) concluíram que o algoritmo NCL trata o comitê de redes neurais artificiais como um macro modelo e por isso apresenta problemas com a regularização. Para resolver este problema propuseram o RNCL – *Regularized Negative Correlated Learning*.

Recentemente Hadavandi *et al.* (2015) propuseram o algoritmo denominado *sequential negative correlated learning* (SNCL), que tem como principal característica a introdução de um termo de penalidade na função custo do modelo que está sendo gerado que depende da correlação com os modelos gerados anteriormente, quanto maior a correlação maior é a penalidade.

2.3.2 Seleção de componentes

É possível que nem todos os componentes gerados contribuam para a melhoria do desempenho de generalização do *ensemble*, sendo necessária uma seleção dos componentes de forma a aperfeiçoar sua capacidade de generalização. O QUADRO 3 exibe os principais algoritmos de seleção encontrados na literatura.

2.3.2.1 Exponencial

A seleção exponencial se refere a testar todas as combinações dos modelos disponíveis, daí o seu nome, quanto maior a quantidade de modelos o esforço computacional aumenta exponencialmente, por isso não existem muitos trabalhos na literatura que abordam este método, mas Aksela (2003) propõe o cálculo da correlação dos erros de cada par de modelos e selecionar o subconjunto com menor correlação.

2.3.2.2 Aleatória

Nesta categoria entra a aplicação de algoritmos genéticos, busca Tabu e aprendizado incremental baseado em população utilizados por Ruta e Gabrys (2001). Zhou *et al.* (2002) apresentam uma abordagem denominada GASEN (*Genetic Algorithm based Selective Neural Network Ensemble*) para a seleção de *ensembles* que primeiramente treina uma quantidade de redes neurais artificiais e atribui um peso aleatório com distribuição uniforme a cada uma delas e evolui os

pesos de forma que o *ensemble* de todas as redes consiga aproximar o resultado esperado, quando finalmente são selecionados os componentes efetivos para compor o *ensemble* final com base nos pesos evoluídos.

ALGORITMO	PUBLICAÇÃO	CATEGORIA	
Correlação máxima entre pares de preditores	Aksela (2003)	Exponencial	
Algoritmo genético, Tabu e aprendizado incremental baseado em população	Ruta e Gabrys (2001)	Aleatória	
GASEN - <i>Genetic Algorithm based Selective Neural Network Ensemble</i>	Zhou <i>et al.</i> (2002)		
<i>Greedy hill-climbing</i>	Caruana <i>et al.</i> (2004)		
Algoritmos imuno-artificiais	Castro <i>et al.</i> (2005) Zhang <i>et al.</i> (2005)		
RSE	Li e Zhou (2009)		
Algoritmos genéticos para criação de matriz de divergências	Cavalcanti <i>et al.</i> (2016)		
FSS	Perrone e Cooper (1993)	Seleção para frente	
Cw/oE	Coelho e Von Zuben (2006)		
CwE			
<i>Max-linear Relevance and Max-linear complementarity</i>	Che (2015)		
Pw/oE	Coelho e Von Zuben (2006)	Seleção para trás	
PwE			
<i>Reduce error pruning with backfitting</i>	Margineantu e Dietterich (1997)	Seleção para trás e para frente misturada	
K+1 modelos	Moreira <i>et al.</i> (2006)		
Limite de correlação	Perrone e Cooper (1993)	Poda por ranking	
Precisão	Partridge e Yates (1996)		
Precisão + Diversidade	Rooney <i>et al.</i> (2004)		
Diagrama kappa-erro	Margineantu e Dietterich (1997)		
Complementaridade	Martínez-Muñoz e Suárez (2004)		
Ordenação apropriada	Martínez-Muñoz e Suárez (2006)		
Baseado no <i>boosting</i>	Martínez-Muñoz e Suárez (2007)		
Ponderação baseada no ranking	Adhikari <i>et al.</i> (2015)		
ISCG-Ranking	Ykhlef e Bouchaffra (2015)		
Dados desbalanceados	Galar <i>et al.</i> (2016)		
On-line com fator de esquecimento	Soares e Araújo (2016)		
Mapa auto organizado	Fonseca e Gómez (2016)		
Aglomerativo hierárquico	Giacinto <i>et al.</i> (2000)		Agrupamento de dados
<i>k-means</i>	Lazarevic e Obradovic (2001)		
Recozimento determinístico	Bakker e Heskes (2003)		
ARIA	Coelho e Von Zuben (2006)		

QUADRO 3 - ALGORITMOS DE SELEÇÃO DE COMPONENTES PARA COMITÊS

FONTE: O autor (2016)

Nesta linha de algoritmos aleatórios pode-se citar Caruana *et al.* (2004) que utilizaram o algoritmo *Greedy Hill-Climbing*, e Castro *et al.* (2005) bem como Zhang *et al.* (2005) que aplicaram algoritmos de imunidade artificial. É possível ainda citar Li e Zhou (2009) que propuseram um método que converte o problema de seleção em um problema de otimização quadrática denominado RSE.

Recentemente, Cavalcanti *et al.* (2016) propuseram um método de seleção que utiliza algoritmos genéticos para criar matriz de divergências entre os modelos, são então formados grafos que indicam quais modelos devem compor o comitê.

2.3.2.3 Seleção para frente

A seleção para frente consiste em iniciar com um comitê vazio e iterativamente adicionar modelos com o objetivo de melhorar o desempenho. Perrone e Cooper (1993) e Coelho e Von Zuben (2006) utilizam dois métodos de seleção para frente. O primeiro denominado Cw/oE (*constructive without exploration*) e o segundo de CwE (*constructive with exploration*), sendo que no primeiro os modelos são testados e classificados em um *ranking*, o comitê adiciona modelos por ordem de classificação no *ranking* até que não haja mais melhora no desempenho, já no segundo método todos os modelos são testados no comitê a cada nova iteração e aquele que melhora seu desempenho é incluído.

Che (2015) utiliza a informação mútua para estabelecer métricas de desempenho que consideram a maximização da divergência entre os modelos e a dependência entre entradas e saídas do comitê, denominadas *max-linear-relevance* e *min-linear-redundancy*, a partir das quais os melhores modelos são sequencialmente inseridos no comitê.

2.3.2.4 Seleção para trás

De forma inversa à seleção para frente, na seleção para trás o comitê é iniciado com todos os modelos e iterativamente se remove um modelo que é responsável pela degradação do desempenho do comitê.

Coelho e Von Zuben (2006) utilizam dois métodos de seleção para trás. O primeiro denominado Pw/oE (*pruning without exploration*) e o segundo de PwE (*pruning with exploration*), sendo que no primeiro os modelos são testados e classificados em um *ranking*, o comitê remove os piores modelos por ordem de classificação no *ranking* até que não haja mais melhora no desempenho, já no segundo método todos os modelos são testados no comitê a cada nova iteração e aquele que ao ser removido melhora seu desempenho é excluído.

2.3.2.5 Seleção para trás e para frente misturada

Margineantu e Dietterich (1997) propuseram um algoritmo que inicia com a seleção para frente para os três primeiros modelos adicionados ao comitê, quando então efetua alternadamente uma seleção para trás e outra para frente.

Já o algoritmo $k + 1$ modelos, proposto por Moreira *et al.* (2006), inicia o comitê com um número de modelos pré-determinado k e executa alternadamente uma seleção para frente, obtendo $k + 1$ modelos e outra para trás, retornando a k modelos. O processo termina quando o mesmo modelo é selecionado nos dois passos.

2.3.2.6 Poda por ranking

Esta classe de algoritmos de seleção ordena os modelos de acordo com alguma métrica e gera um comitê a partir dos melhores modelos, definidos com um valor de desempenho limite.

Perrone e Cooper (1993) cria uma matriz de correlação entre os modelos candidatos e primeiramente elimina todos aqueles que são fortemente correlacionados de acordo com um limite pré-determinado.

Uma abordagem direta da poda por ranking é utilizada em Partridge e Yates (1996), onde é calculada a precisão de cada modelo, formado um ranking e escolhidos os melhores para compor o comitê. O problema é que apenas o critério

de precisão não garante diversidade, isto motivou Rooney *et al.* (2004) a usar o mesmo método, porém calculando o desempenho com uma métrica que equilibra precisão e diversidade.

Outro algoritmo que forma um ranking por precisão e divergência é o diagrama kappa-erro (Margineantu e Dietterich, 1997), no qual os modelos são pontos que representam uma medida de precisão e uma de divergência, no diagrama é definida uma curva convexa que separa os modelos aceitáveis dos outros.

Martínez-Munõz e Suarez (2004) propõem a seleção por complementaridade. Já em Martínez-Munõz e Suarez (2006) é proposto o método de ordenação apropriada, ou seja, os modelos possuem uma ordenação ótima que ao ser obedecida é possível encontrar a quantidade ótima de modelos a serem adicionados ao comitê. Ainda os autores Martínez-Munõz e Suarez (2007) propõem a ordenação dos modelos usando o algoritmo Adaboost (Freund e Schapire, 1996).

Pode-se dizer que o estágio de seleção é um caso particular do estágio de combinação dos componentes do comitê, quando os pesos alocados para um ou mais componentes é nulo. Em Adhikari *et al.* (2015) os modelos são classificados em um ranking e apenas os melhores são selecionados como membros do comitê. Ykhlef e Bouchaffra (2015) propuseram um método de seleção baseado em um jogo subgráfico induzido que cria um ranking através de uma métrica denominada valor de Shapley.

Um problema comum que pode surgir ao se tratar de problemas reais é o desequilíbrio de dados, ou seja, podem existir mais exemplos para uma pequena região de operação do sistema do que em outras que induzem uma polarização para esta região mais densa de exemplos. Para abordar este problema Galar *et al.* (2016) propuseram um esquema de poda baseada em ordenação que compensa este desequilíbrio de dados.

A abordagem de comitês em tarefas de previsão on-line é abordada por Soares e Araújo (2016). Geralmente o processo de construção de um comitê é realizado uma vez e depois é aplicado sobre novos conjuntos de dados quando necessário, já em aplicações on-line deseja-se aplicar o comitê a cada instante e continuamente, e conseqüentemente a medida que o tempo passa as características

do processo podem ser alteradas, exigindo que o modelo seja ajustado à nova realidade. No âmbito dos comitês os autores propuseram um método para atualizar a seleção dos membros do comitê em aplicações on-line baseado em poda por ranking.

Ao utilizar comitês de redes neurais artificiais de mapas auto organizados (*self organized maps* - SOM) para previsão de séries temporais, Fonseca e Gómez (2016) propuseram um método de poda que utiliza a distância entre neurônios no mapa para definir métricas de precisão e diversidade entre os modelos preditores e selecionar apenas a mais representativa de cada grupo para compor o comitê.

2.3.2.7 Agrupamento de dados (*clustering*)

A ideia central da seleção por agrupamento é agrupar os modelos e tomar indivíduos representativos de cada grupo para compor o comitê. Em Lazarevic e Obradovic (2001) o algoritmo *k-means* é usado para agrupar os vetores de previsão de cada modelo em k grupos, dos quais é escolhido um modelo representativo de cada e, na sequência, são ordenados de acordo com algum critério e adicionados ao comitê até que o desempenho comece a degradar.

De forma similar, Giacinto *et al.* (2000), Bakker e Heskes (2003) e Coelho e Von Zuben (2006) alteram apenas o algoritmo de agrupamento de dados para o Aglomerativo Hierárquico, o Recozimento Determinístico e o ARIA – *Adaptive Radius Immune Algorithm*, respectivamente.

2.3.3 Combinação de componentes

Esta é a etapa na qual as previsões dos modelos selecionados são integradas para formar a saída do comitê. O QUADRO 4 exhibe os principais algoritmos desta etapa.

ALGORITMO	PUBLICAÇÃO	CATEGORIA
Generalização empilhada	Wolpert (1992)	Ponderação constante
BEM	Perrone e Cooper (1993)	
GEM		
Regressão empilhada	Breiman (1996)	
PCR*	Merz e Pazzani (1999)	
Seleção com reposição	Caruana <i>et al.</i> (2004)	
Algoritmos evolucionários	Ortiz-Boyer <i>et al.</i> (2005)	
Mediana	Barrow <i>et al.</i> (2010)	
Moda	Kourentzes <i>et al.</i> (2014)	
Minimização do <i>MSFE</i>	Conflitti <i>et al.</i> (2015)	
Média truncada <i>Bayesian Model Averaging (BMA)</i>	Hassan <i>et al.</i> (2015)	
Meta árvore de decisão	Todorovski e Dzeroski (2003)	Ponderação não constante: áreas de especialização
SVM com <i>kernel</i> Gaussiano	Yankov <i>et al.</i> (2006)	
Meta-heurística para generalização empilhada	Shunmugapriya e Kanmani (2013)	
<i>First work</i>	Bates e Granger (1969)	Ponderação não constante: abordagem dinâmica
Ponderação baseada na variância	Tresp e Taniguchi (1995)	
Matriz de desempenho para avaliar modelos locais	Merz e Pazzani (1996)	
k-vizinhos mais próximos ponderado	Puuronen <i>et al.</i> (1999)	
Inversamente proporcional	Wang <i>et al.</i> (2003)	
DWS	Rooney <i>et al.</i> (2004)	

QUADRO 4 - ALGORITMOS DE COMBINAÇÃO DE COMPONENTES PARA COMITÊS

FONTE: O autor (2016)

2.3.3.1 Ponderação constante

Wolpert (1992) apresenta o método da generalização empilhada (*stacked generalization*) para a combinação de componentes, no qual são criados modelos preditores individuais assim como nos outros métodos, contudo as suas saídas em conjunto com as saídas desejadas são utilizadas para criar um novo conjunto de dados de treinamento, o qual é utilizado para estimar um meta-modelo que mapeia as saídas dos preditores na saída desejada.

Em Perrone e Cooper (1993) a média simples é apresentada como um método básico de combinação para *ensembles* (*Basic Ensemble Method* – BEM) e a média ponderada como um método de combinação generalizado (*Generalized Ensemble Method* – GEM).

O algoritmo da regressão empilhada proposto por Breiman (1996) é baseado na generalização empilhada, a qual foi inicialmente proposta para problemas de classificação, e adapta as equações do algoritmo para problemas de regressão. A partir deste algoritmo, Ortiz-Boyer *et al.* (2005) utilizaram algoritmos evolucionários para determinação dos coeficientes para ponderar cada modelo na combinação.

Quando se tem muitos modelos candidatos para um comitê é mais provável que se perceba a multicolinearidade, ou seja, que existam modelos muito parecidos e não seja vantajoso incluí-los simultaneamente no comitê. Para abordar este problema Merz e Pazzani (1999) propuseram o algoritmo PCR*.

A reamostragem com reposição sobre o conjunto de modelos ao invés do conjunto de dados foi utilizada por Caruana *et al.* (2004). Kourentzes *et al.* (2014) propõem a utilização da moda para combinação de componentes do *ensemble* através da estimação de densidade de kernel (*kernel based density*), que não exige a discretização dos valores. Segundo os autores, seus resultados indicam que a moda supera a média e a mediana porque é insensível a *outliers* e desvios da normalidade.

Em Barrow *et al.* (2010) são comparados *ensembles* criados pela combinação através da média simples e da mediana, e conclui que a última apresenta melhores resultados.

Mais recentemente Hassan *et al.* (2015) examinaram a eficiência de três algoritmos de combinação para previsões de 100 redes neurais artificiais sobre dados de demanda amostrados a cada 30 minutos na Austrália e em Nova York. As abordagens utilizadas foram a média simples, a média truncada (*trimmed mean*) e a ponderação baseada em modelo Bayesiano denominada BMA (*Bayesian Model Averaging*). A primeira conclusão que chegaram é que o *ensemble* se saiu melhor do que muitos modelos individuais e bem melhor do que o modelo inocente, a segunda conclusão é que a média truncada resultou em desempenho inferior com relação aos outros dois métodos.

Em busca de uma combinação ótima das previsões Conflitti *et al.* (2015) propuseram a minimização do índice MSFE (*mean square forecast error*) para estabelecer os pesos de combinação para cada preditor.

2.3.3.2 Ponderação não constante: áreas de especialização

Esta abordagem considera modelos que não tentam resolver o problema diretamente, mas recomendam quais os melhores modelos de um conjunto são os mais indicados para cada entrada, também conhecido como mistura de especialistas.

Todorovski e Dzeroski (2003) propõem uma meta árvore de decisão para problemas de classificação, mas que pode ser estendida para problemas de regressão. Já Yankov *et al.* (2006) utilizam SVM com *kernel* Gaussiano para escolher entre dois modelos.

Em Shunmugapriya e Kanmani (2013) é proposta uma meta-heurística para seleção dos melhores modelos para compor as camadas de nível-0 e de nível-1 em *ensembles* de generalização empilhada através do algoritmo da colônia artificial de abelhas ABC (*Artificial Bee Colony*), que foi aplicado em problemas de classificação do UCI (*University of California Irvine*).

2.3.3.3 Ponderação não constante: abordagem dinâmica

Nesta categoria os modelos também são alocados de acordo com o conjunto de entradas e no histórico de desempenho do preditor, contudo esta decisão é realizada a cada padrão de entrada apresentado enquanto que na abordagem por áreas de especialização esta decisão é realizada previamente.

O primeiro trabalho ao utilizar este método foi realizado por Bates e Granger (1969) que propuseram um algoritmo que decidia entre dois modelos preditores com base em valores de erro passados. Na mesma linha Tresp e Taniguchi (1995) propõem a ponderação dos modelos baseada na variância dos erros passados e Wang *et al.* (2003) utiliza pesos inversamente proporcionais ao erro esperado de cada modelo, se aproximando da abordagem baseada na variância.

Merz e Pazzani (1996) propõe a criação de uma matriz de desempenho de cada modelo relacionando um índice para cada exemplo passado. Puuronen *et al.* (1999) utilizam o algoritmo k-vizinhos mais próximos ponderado para decisão do

preditor enquanto Rooney *et al.* (2004) adapta o algoritmo k-vizinhos mais próximos ponderado para um método denominado DWS, *Dynamic Weighting with Selection*.

2.4 PRÉ-PROCESSAMENTO DOS DADOS

Em Alexandridis e Zapranis (2014) é apresentado um caso de modelagem da velocidade média diária do vento em Nova York em que o pré-processamento dos dados é realizado através da transformação Box-Cox, a qual preservou a característica de periodicidade dos dados e aproximou os dados da distribuição normal.

Segundo Graff *et al.* (2014) a transformada Box-Cox corresponde à tendência existente em um conjunto de dados e Villanueva (2006), bem como Brockwell e Davis (2002) e Anderson-Sprecher (1994), expõem que as tendências tornam as séries temporais não estacionárias, sendo necessário diferenciá-las o quanto for necessário a fim de se remover esta tendência, contudo a diferenciação pode desestabilizar a variância e não corrigir a assimetria da distribuição, sendo recomendada a utilização da transformação Box-Cox.

3 MATERIAIS E MÉTODOS

Este capítulo é dedicado à descrição dos materiais e métodos empregados nos experimentos do próximo capítulo. Inicialmente são descritos os conjuntos de dados de aferição e o seu pré-processamento, logo após são descritos os métodos para definição da janela de predição e para seleção de variáveis, na sequência é descrita a estimação das *wavenets*, são apresentados os métodos de previsão utilizados bem como os índices de avaliação do desempenho. Por fim são abordados os métodos de aprendizado de comitês e a organização de alguns conceitos úteis relacionados ao nosso estudo.

3.1 PROCEDIMENTOS METODOLÓGICOS

Esta seção apresenta a descrição, a caracterização e as fases da pesquisa, as limitações do estudo e a organização desta dissertação. A FIGURA 7 apresenta um fluxograma da proposta de pesquisa onde é possível observar que a partir de uma série de valores horários de demanda é realizado um pré-processamento, na sequência são formados um conjunto de dados para treinamento e outro para teste. Os dados de treinamento são utilizados para estimação dos modelos individuais e, logo após, para aprendizado dos comitês. Uma vez produzidos os comitês é realizada a previsão de curto prazo com o conjunto de dados de teste, este processo é realizado iterativamente em um esquema de validação cruzada.

Esta pesquisa tem caráter aplicado, já que os experimentos são realizados com dados reais, de forma quantitativa ao se calcular índices de desempenho para cada modelo preditor e exploratória porque procura conhecer o comportamento dos comitês de *wavenets* na classe de problemas de previsão de demanda.

Este trabalho está limitado quanto a três variáveis principais, que são os modelos sujeitos a comparação, os métodos de aprendizado de comitês e aos problemas aplicados. Foram escolhidos alguns modelos comumente encontrados na literatura para aplicação na previsão de curto prazo dentre muitas opções existentes, do mesmo modo para o aprendizado dos comitês foram utilizados os métodos ditos

convencionais e os problemas experimentados são limitados a duas séries horárias de demanda.

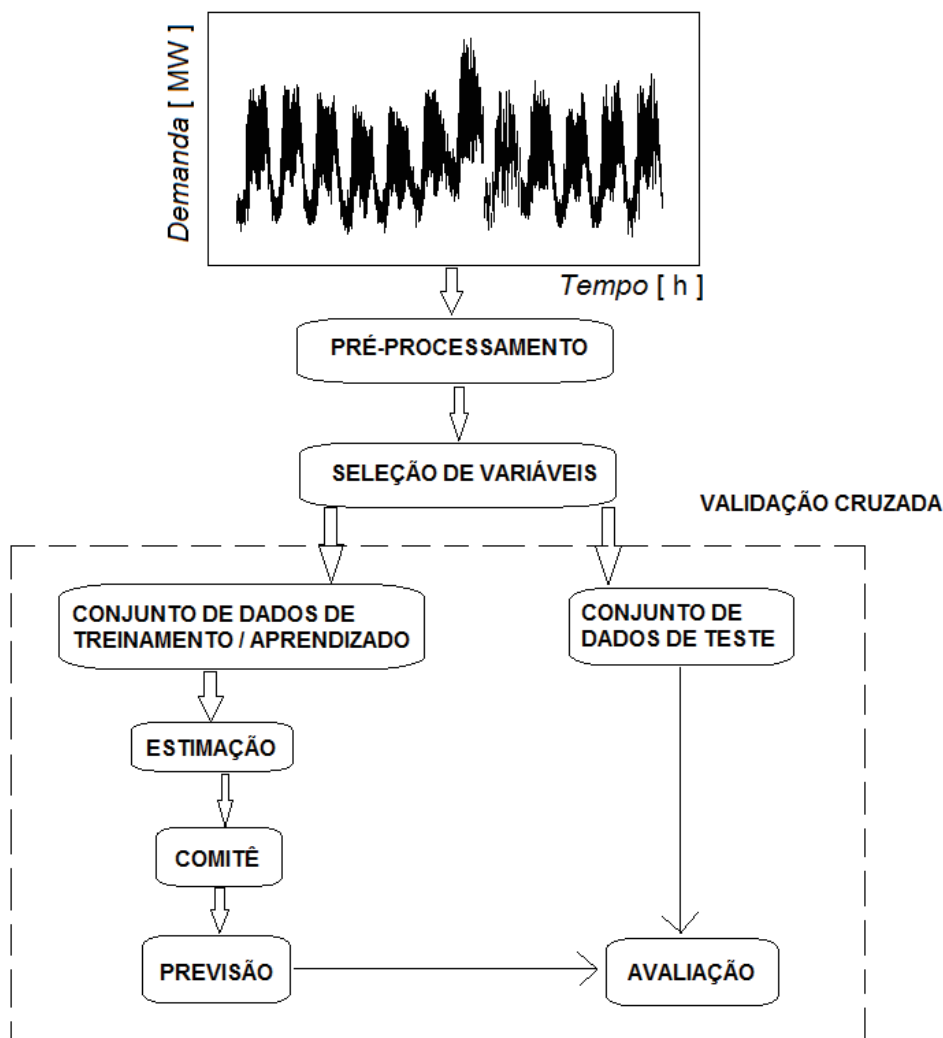


FIGURA 7 – PROPOSTA DE PESQUISA

FONTE: O autor (2016)

Na FIGURA 7, para o pré-processamento é utilizada a transformação Box-Cox, um teste de não linearidade, o cálculo da janela de predição e a conversão da série temporal em dados de entrada e saída. Para a seleção de variáveis é utilizado um método de seleção para frente. A partir do conjunto de dados e saídas processado e com variáveis selecionadas é realizada a partição dos dados em conjuntos de treinamento, o qual contempla 90% dos dados, e de teste, que contempla os outros 10%. O conjunto de treinamento é utilizado para estimação das

wavenets através do ambiente computacional Matlab. Os comitês são criados a partir dos métodos de reamostragem com reposição, validação cruzada, dizimação de entradas, seleção construtiva, combinação pela média simples, moda, mediana e generalização empilhada. Com os comitês criados obtém-se sua previsão sobre o conjunto de testes utilizando a estratégia iterativa, a qual é avaliada com o coeficiente de correlação múltipla. Os procedimentos no interior da caixa delimitada pela linha tracejada são repetidos em um esquema de validação cruzada com dez repetições, de forma que o algoritmo é testado em todos os dados disponíveis.

3.2 PRÉ-PROCESSAMENTO DOS DADOS

Nesta pesquisa são utilizados dois conjuntos de dados, sendo o primeiro do repositório Kaggle, referente à competição de previsões para demanda horária no período de uma semana de 20 zonas de consumidores nos Estados Unidos chamada GEFCom 2012 (*Global Energy Forecasting Competition 2012*) publicado por Hong *et al.* (2014) e o segundo da Rede Europeia de Operadores de Sistemas de Transmissão de Eletricidade (ENTSO-E, *European Network of Transmission System Operators for Electricity*) referente a demandas horárias da Itália no ano de 2015.

3.2.1 GEFCom 2012

Este conjunto de dados provém de 20 zonas de consumo americanas com demandas horárias de energia elétrica em kW, no período de 06/03/2005 a 07/07/2008, com lacunas a cada intervalo de 3 meses aproximadamente, as quais devem ser previstas. Juntamente com os dados de demanda estão os dados de temperatura para serem usados no espaço de características, e, além disso, os valores ausentes no conjunto de estimação estão em um conjunto separado para testes e validação. Este trabalho utiliza a série apenas no período de 2005 e a zona de consumo nº 1.

3.2.2 Itália

Em 2009 a União Europeia criou a rede europeia dos operadores do sistema de transmissão, denominada ENTSO-E (*European Network of Transmission System Operators*), para otimizar seu mercado interno de energia, a qual deu transparência aos dados históricos de demanda horária de energia elétrica de 35 países membros.

Para testar o desempenho dos algoritmos neste trabalho foi obtido o histórico de demanda no período de dezembro de 2014 a dezembro de 2015 na Itália. Os dados são importados para o ambiente computacional Matlab a partir de planilhas eletrônicas mensais e manipulados de forma a compor um vetor com a série histórica.

3.2.3 Transformação Box-Cox

As séries temporais de demanda de energia são pré-processadas através da transformada Box-Cox com o objetivo de remover a tendência e aproximá-la da distribuição estatística normal, e é dada por

$$dados(\lambda) = \begin{cases} \frac{dados^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(\lambda), & \lambda = 0 \end{cases} \quad (3.1)$$

onde λ é o parâmetro estimado pela maximização da log-verossimilhança e $dados(\lambda)$ representa os valores da série original após a transformação.

3.2.4 Teste de não linearidade

Antes de empregar um modelo não linear para a modelagem de um sistema é importante descobrir, a partir do conjunto de dados disponíveis, se o sistema exhibe características não lineares, e para isso é necessário aplicar um teste de não linearidade.

Um teste proposto por Billings e Voon (1983) estabelece que a correlação cruzada entre os valores da série obedecem a

$$\begin{cases} \phi_{y'y'^2}(\tau) = 0, (\tau = 0, 1, \dots) \Rightarrow \text{sistema linear} \\ \phi_{y'y'^2}(\tau) \neq 0, (\tau = 0, 1, \dots) \Rightarrow \text{sistema não linear} \end{cases} \quad (3.2)$$

onde k é a amostra de teste, \bar{y} é a média dos valores da série $y(k)$, e $y'(k) = y(k) - \bar{y}$.

Na prática considera-se que $\phi_{y'y'^2}(\tau) = 0$ se estiver dentro do intervalo de confiança delimitado por $\pm 1,96\sqrt{N}$, em que N é a quantidade de amostras do conjunto de teste. Mesmo que o sistema seja não linear o valor de $\phi_{y'y'^2}(\tau)$ se aproxima de 0 para um valor de τ suficientemente grande, contudo uma prática adotada é considerar o valor máximo de 5.

3.2.5 Cálculo da janela de predição

Para o treinamento de uma *wavenet* é necessário que se tenha um conjunto de dados de entrada e saída do tipo $[X, y]$ no qual X é uma matriz $N \times d$, onde N é a quantidade de amostras e d a quantidade de variáveis de entrada, ou seja, a dimensão do espaço de entrada, enquanto y é um vetor $N \times 1$ com os valores alvo.

Contudo uma série temporal não possui naturalmente uma configuração de entradas e saídas, sendo necessário escolher quais valores passados serão utilizados para prever o futuro, chamados de regressores. O regressor mais antigo define a janela de predição.

Billings (2013) propõe a utilização da taxa de redução de erros (ERR – *Error Rate Reduction*) para o cálculo da janela de predição. Para cada variável, neste caso um valor anterior da série, também denominado regressor ou atraso, será calculado o seu ERR, na sequência são ordenados em ordem decrescente e realizada a soma dos ERR individuais, ou SERR, até que se atinja o limite de tolerância adotado. Neste caso foi adotado o valor em que não há mais aumento da SERR.

Para este cálculo utiliza-se um modelo linear porque se entende que caso o sistema modelado seja não linear, a quantidade de regressores necessária para representá-lo com um modelo não linear, que é o caso de uma *wavenet*, será menor, portanto o sistema é representado por,

$$y(k) = \sum_{i=1}^M g_i w_i(k) + e(k) \quad (3.3)$$

onde M é a quantidade de regressores candidatos, escolhida suficientemente grande para contemplar o regressor mais antigo que não terá mais influência na variável de saída, adotada neste trabalho como 500, $w_i(k)$ são os regressores ortogonalizados, g_i são os parâmetros do modelo e $e(k)$ o erro do modelo.

Quando se diz que os regressores são ortogonais, quer-se dizer que satisfazem as seguintes equações,

$$\sum_{k=1}^N w_i(k)w_j(k) = \begin{cases} d_i, & i = j \\ 0, & i \neq j \end{cases} \quad (3.4)$$

$$d_i = \sum_{k=1}^N w_i^2(k) \neq 0 \quad (3.5)$$

A partir destas equações aplica-se o procedimento de ortogonalização

$$\begin{cases} w_1(k) = p_1(k) \\ w_2(k) = p_2(k) - a_{1,2}w_1(k) \\ w_3(k) = p_3(k) - a_{1,3}w_1(k) - a_{2,3}w_2(k) \\ \vdots \\ \vdots \\ w_m(k) = p_m(k) - \sum_{r=1}^{m-1} a_{r,m}w_r(k), m = 2, 3, \dots, M \end{cases} \quad (3.6)$$

onde $p_m(k)$ é o m -ésimo regressor antes da ortogonalização na amostra k . Os valores $a_{r,m}$ são determinados por

$$a_{r,m} = \frac{\sum_{k=1}^N p_m(k)w_r(k)}{\sum_{k=1}^N w_r^2(k)}, 1 \leq r \leq m - 1 \quad (3.7)$$

Os parâmetros g_i são

$$g_i = \frac{\sum_{k=1}^N y(k)w_i(k)}{\sum_{k=1}^N w_i^2(k)}, i = 1, 2, \dots, M \quad (3.8)$$

Dadas as equações anteriores, a taxa de redução de erro para cada variável i é

$$ERR_i = \frac{g_i^2 \langle \mathbf{w}_i, \mathbf{w}_i \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \cdot 100\% \quad (3.9)$$

3.2.6 Conversão da série temporal em dados de entrada e saída

O treinamento das *wavenets* exige, em muitos casos, a manipulação do conjunto de observações de entrada em operações como a reamostragem aleatória com reposição, por exemplo, e como se trata de séries temporais esta manipulação retira o sentido sequencial da série.

Para lidar com esta situação adota-se o seguinte procedimento, na FIGURA 8 a série é composta de N amostras e a janela de atrasos foi definida a título de ilustração em 6, ou seja, para a previsão do próximo valor da série são necessários 6 valores passados, sendo cada coluna considerada como uma entrada, ou um regressor. A janela de atrasos é determinada de acordo com a periodicidade de repetição do padrão da série, a qual no caso da previsão de demanda de energia elétrica horária de curto prazo é definida como 168, ou seja, possui um padrão de repetição semanal, que corresponde a 168 regressores para estimar a demanda na próxima hora.

Esta grande quantidade de entradas pode causar preocupação porque aumenta drasticamente o tempo de processamento de uma *wavenet*, contudo será percebido que a quantidade de regressores necessária é bem menor.

A quantidade de pares (\vec{X}, Y) de entrada e saída é limitada ao tamanho da série temporal subtraído de 168, que é a janela de atrasos de uma semana. Além disso, são separadas 168 amostras para validação, que não podem estar contidas no conjunto de treinamento, significando que o tamanho total do conjunto de treinamento é o tamanho total da série subtraído de 336 amostras.

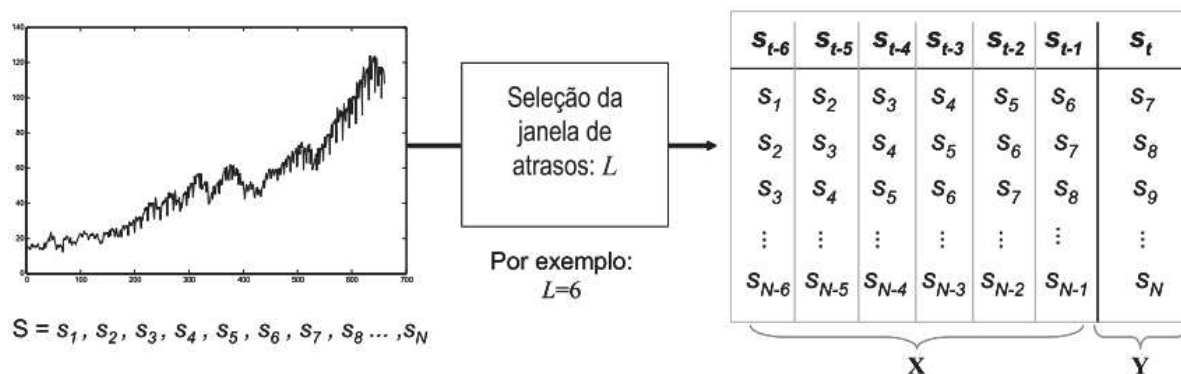


FIGURA 8 - PREPARAÇÃO DOS DADOS DE TREINAMENTO PARA O *BAGGING*

FONTE: Villanueva (2006)

3.3 SELEÇÃO DE VARIÁVEIS

Quando a quantidade de variáveis é da ordem de dezenas ou centenas o esforço computacional para estimação dos modelos pode se tornar impraticável, além disso, os modelos tendem a ser polarizados perdendo capacidade de generalização, por isso é necessário selecionar dentre todas as variáveis apenas aquelas que são mais representativas, ou seja, que captam a maior parte da dinâmica do sistema que está sendo modelado.

Utiliza-se o algoritmo de seleção sequencial para frente, ou *sequential forward selection* (SFS). Este algoritmo começa com um conjunto de entradas vazio, adiciona variáveis sequencialmente e avalia o desempenho do modelo com a variável adicionada até que se atinja um critério de parada. Inicialmente cada variável é incluída isoladamente no conjunto de entrada, para cada uma é estimada uma *wavenet* utilizando-se 80% das amostras e avaliando o seu desempenho nos 20% restantes através do coeficiente de correlação múltipla (R^2) sobre a previsão um passo à frente. Em cada iteração é realizada a validação cruzada com cinco partições.

3.4 ESTIMAÇÃO DAS WAVENETS

As *wavenets* são estimadas através da caixa de ferramentas para identificação de sistemas (*Systems Identification Toolbox* – SITB) do ambiente computacional Matlab, que utiliza a equação

$$\begin{aligned}
 F(x) = & (x - r)PL + a_{s,1}f(b_{s,1}((x - r)Q - c_{s,1})) + \dots \\
 & + a_{s,ns}f(b_{s,ns}((x - r)Q - c_{s,ns})) \\
 & + a_{w,1}g(b_{w,1}((x - r)Q - c_{w,1})) + \dots \quad (3.10) \\
 & + a_{w,nw}g(b_{w,nw}((x - r)Q - c_{w,nw})) + d
 \end{aligned}$$

Nesta equação o primeiro termo é a parte linear, o último termo é o *bias*, as funções $a_w g(b_w((x - r)Q - c_w))$ são as wavelets e a_w, b_w e c_w os seus parâmetros. As funções $a_s f(b_s((x - r)Q - c_s))$ são funções de escala, que não aparecem originalmente na estrutura das *wavenets*, inseridas para estabilizar oscilações indesejadas aumentando a regularização do estimador. Os demais parâmetros são:

x : vetor linha m -dimensional de entrada (regressores);

r : vetor $(1 \times m)$ com o valor médio de cada regressor;

P : matriz $(m \times p)$ de projeção da análise de componentes principais dos regressores lineares;

Q : matriz $(m \times q)$ de projeção da análise de componentes principais dos regressores não lineares;

L : vetor $(p \times 1)$ com os coeficientes dos pesos das conexões diretas entre as entradas e o neurônio de saída que modela a parte linear;

a_s, b_s, c_s : coeficientes de escala, dilatação e translação respectivamente para a função de escala;

a_w, b_w, c_w : coeficientes da wavelet, dilatação e translação respectivamente para as wavelets;

d : polarização ou *bias*.

A dimensão da matriz P é $(m \times p)$, sendo m a quantidade de regressores e $p < m$ se existirem regressores linearmente dependentes ou $p = m$ caso todos os regressores sejam linearmente independentes, como consequência direta da análise de componentes principais (*Principal Component Analysis – PCA*). Já para a matriz Q , cuja dimensão é $(m \times q)$, $m = q$, ou seja, q recebe o mesmo valor da quantidade de regressores m .

A função de escala utilizada é

$$f(x) = e^{-\frac{1}{2}x'x} \quad (1)$$

A *wavelet* utilizada é a do tipo chapéu mexicano

$$g(x) = (1 - x'x)e^{-\frac{1}{2}x'x} \quad (2)$$

Como visto anteriormente no capítulo 2 a análise de *wavelet* produz caixas de Heisenberg no plano tempo-frequência, as quais também são denominadas células, o Matlab determina automaticamente a quantidade de células e a resolução de cada uma. O foco da nossa estimação é a previsão e não a simulação, por isso a função custo otimizada durante o treinamento da *wavenet* é a norma-2 de $y - y_{pred}$. Os métodos de busca utilizados são o Gauss-Newton, o Gauss-Newton adaptativo e o Mínimos Quadrados Não linear abordados por Nocedal e Wright (2013), Levenberg-Marquardt (Levenberg, 1944), Gradiente descendente (Petrova e Solov'ev, 1997).

3.5 MÉTODO DE PREVISÃO

Ben Taieb *et al.* (2012) revisa as estratégias existentes para a previsão multi-passos à frente e define o problema da seguinte forma. Dada uma série $[y_1, \dots, y_N]$ que possui N valores, a tarefa consiste em prever os próximos H valores da série $[y_{N+1}, \dots, y_{N+H}]$ em que H é denotado horizonte de previsão. Neste trabalho o horizonte de previsão considerado é de um até 168.

A estratégia empregada, denominada iterativa ou recursiva, consiste em treinar uma *wavenet* denotada por f para a previsão um passo à frente, tal que

$$y_{t+1} = f(y_t, \dots, y_{t-d+1}) + e_t \quad (3.11)$$

onde t é o instante de tempo, d é o atraso máximo e e_t é o erro do modelo.

A partir da equação anterior obtém-se a previsão um passo à frente, este valor previsto é adicionado ao vetor de entradas utilizado na próxima iteração e assim sucessivamente até que se complete todo o horizonte de previsão.

Uma característica da previsão iterativa é o acúmulo dos erros quando $H > 1$ porque os valores de entrada não são mais os valores medidos como na primeira iteração, e na medida em que H aumenta o erro acumulado tende a aumentar até um ponto em que a previsão não está mais correlacionada com a série.

3.6 VALIDAÇÃO DOS MODELOS

O coeficiente de correlação múltipla R^2 (Haber e Unbehauen, 1990) dado na seguinte equação é utilizado para validação de modelos, (e.g. Ayala, 2016; Anderson-Sprecher, 1994), o qual é um índice que representa o quanto o modelo estimado representa os dados reais, limitado no intervalo $[0, 1]$ para modelos lineares e no intervalo $(-\infty, 1]$ para modelos não-lineares, sendo que quanto mais próximo de “1” melhor o modelo,

$$R^2 = 1 - \frac{\sum_{t=1}^N |\xi(t)|^2}{\sum_{t=1}^N |y(t) - \bar{y}|^2} \quad (3.12)$$

$$\xi(t) = \hat{y}(t) - y(t) \quad (3.13)$$

onde $\hat{y}(t)$ é a previsão do modelo, $y(t)$ é o valor real do conjunto de teste, $\xi(t)$ é o resíduo da previsão para a amostra t , \bar{y} é a média de y e N é a quantidade de amostras de teste.

Contudo a magnitude dos erros não é o único fator que define a qualidade do modelo, por isso também são utilizados os testes de correlação para séries temporais sugeridos por Billings (2013), que são

$$\begin{cases} \Phi_{\xi' \xi'}(\tau) = \delta(\tau), & \forall \tau \\ \Phi_{\xi' (\xi^2)'}(\tau) = 0, & \forall \tau \\ \Phi_{(\xi^2)' (\xi^2)'}(\tau) = \delta(\tau), & \forall \tau \end{cases} \quad (3.14)$$

em que

$$\begin{cases} \xi'(k) = \xi(k) - \bar{\xi} = \xi(k) - \frac{1}{N} \sum_{k=1}^N \xi(k) \\ (\xi^2)'(k) = \xi^2(k) - \bar{\xi^2} = \xi^2(k) - \frac{1}{N} \sum_{k=1}^N \xi^2(k) \end{cases} \quad (3.15)$$

onde $\xi(k)$ é o resíduo da previsão no instante k . O intervalo de confiança considerado é de aproximadamente 95 %.

3.7 MÉTODOS DE GERAÇÃO DE COMPONENTES PARA O COMITÊ

Nesta seção são apresentados os métodos utilizados para gerar modelos capazes de prever a demanda com pequena polarização e variância e ao mesmo tempo com erros descorrelacionados, os quais ao serem utilizados em um *ensemble* são chamados de componentes ou membros.

3.7.1 Reamostragem com reposição (*bootstrapping*)

A ideia principal do *bootstrapping* (Breiman, 1996) é utilizar conjuntos de treinamento distintos para aumentar a diversidade entre os componentes de um comitê. A FIGURA 9 ilustra sua aplicação, da qual é possível perceber que os “m” conjuntos de dados reamostrados, em que m é a quantidade de WNN's (*wavelet neural networks*) têm o mesmo tamanho N do conjunto original, contudo possuem algumas amostras que se repetem.

Para se utilizar o *bootstrapping* é necessário reorganizar o conjunto de dados de treinamento aleatoriamente. Antes disto, necessita-se do conjunto de dados de treinamento, mais especificamente de valores de entradas e saídas. Contudo, nos problemas de previsão de séries temporais não há sinais de entrada

ou de saída explícitos, por isso a estimação de um modelo considera valores passados da série (regressores) como entradas e o valor atual como saída.

Outro detalhe para a aplicação do *bagging* é a definição do tamanho dos conjuntos de treinamento reamostrados para evitar o sobreajuste e a consequente perda de generalização dos componentes individuais. O valor adotado neste trabalho é de 100%, ou seja, toda amostra do conjunto original tem a mesma chance de estar no novo conjunto.

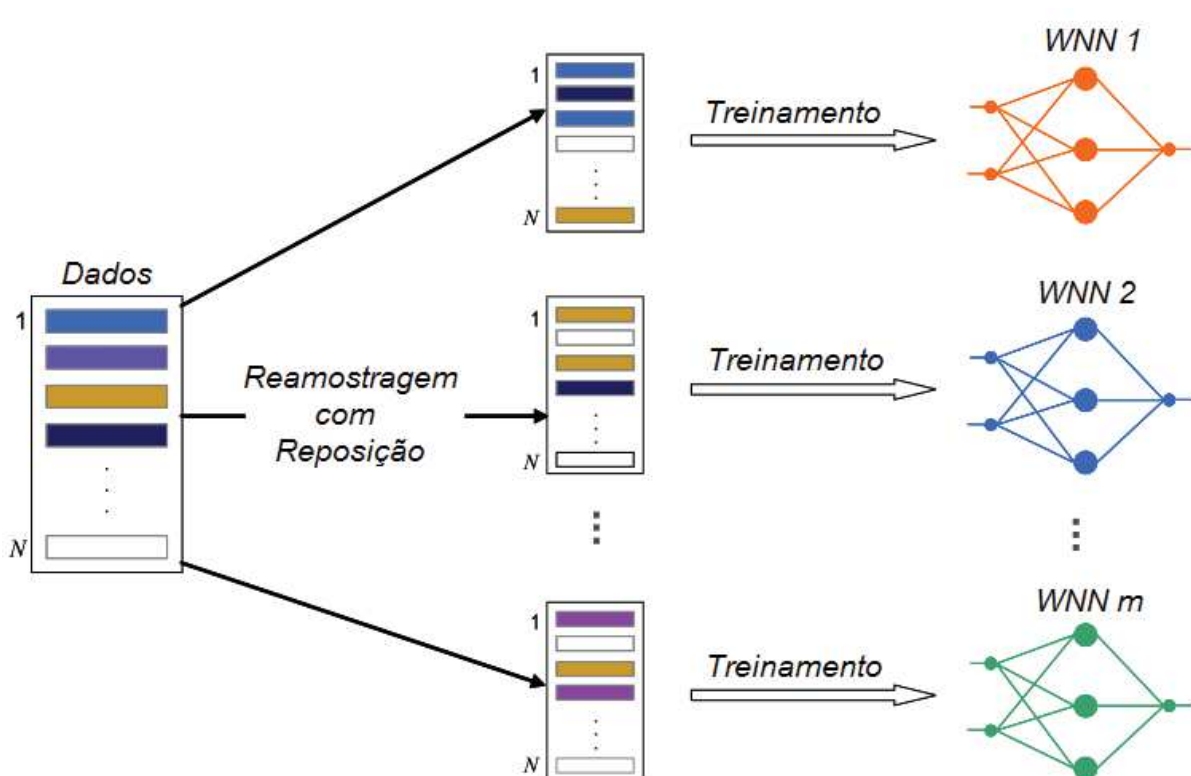


FIGURA 9 - GERAÇÃO DE COMPONENTES VIA *BAGGING*

FONTE: Adaptado de Villanueva (2006)

3.7.2 Validação cruzada (*cross-validation*)

A validação cruzada é um método que gera vários conjuntos de dados para treinamento (*training set*) e teste (*test set*) distintos a partir de um conjunto original e pode ser utilizado para gerar componentes diversos em cada rodada de treinamento (*training round*) na criação de um *ensemble* conforme foi realizado por Wichard e

Ogorzalek (2004), muito útil nos casos em que há escassez de amostras. A FIGURA 10 a seguir ilustra a formação dos diversos conjuntos de dados.

Os autores treinaram diversos tipos de modelos em cada rodada e escolheram aquele modelo que obteve o menor MSE para compor o *ensemble*, seguindo iterativamente até atingir a quantidade de membros desejada para o *ensemble*. Neste trabalho é abordada utilização de *ensemble* de *wavenets*, homogêneo, e por isso não é necessária esta seleção em cada rodada, mesmo assim é esperado que as *wavenets* geradas sejam divergentes porque utilizam conjuntos de treinamento distintos.

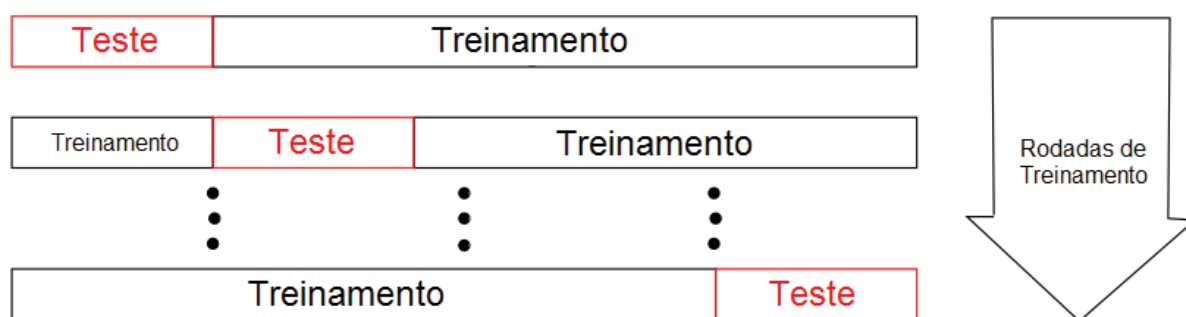


FIGURA 10 – FORMAÇÃO DOS CONJUNTOS DE TREINAMENTO E TESTE

FONTE: Adaptado de Wichard e Ogorzalek (2004)

3.7.3 Dizimação de entradas (*inputs decimation*)

Tumer e Gosh (1996) utilizam a dizimação de entradas para reduzir a correlação entre modelos classificadores, o qual consiste em descartar uma ou mais entradas, reduzindo assim a dimensionalidade do conjunto de treinamento, de modo a treinar cada modelo com um conjunto de dados ligeiramente diferente. A FIGURA 11 ilustra este método.

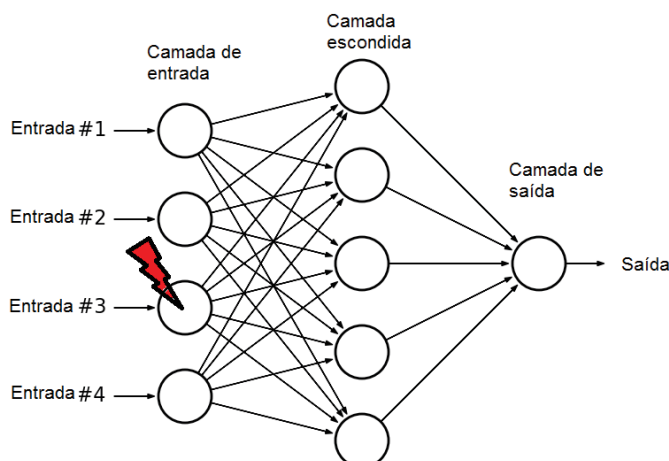


FIGURA 11 - MÉTODO DA DIZIMAÇÃO DE ENTRADAS

FONTE: O autor (2016)

Esta ilustração trata de uma rede neural de uma camada escondida com cinco neurônios, quatro entradas e apenas uma saída, em que a entrada número 3 é dizimada, ou seja, não é considerada na criação do modelo de previsão. Este procedimento é repetido para a quantidade de modelos que se deseja criar e a entrada a ser dizimada é escolhida aleatoriamente com distribuição normal.

3.8 MÉTODO PARA SELEÇÃO DE COMPONENTES PARA O COMITÊ

Perrone e Cooper (1993) propuseram um método para determinação dos pesos ótimos para combinação dos componentes de um *ensemble* através da média ponderada, porém o método utiliza a inversão da matriz de correlação dos erros de previsão dos componentes e por isso pode ser aplicado somente a preditores com garantia de diversidade, caso contrário a matriz de correlação se torna mal condicionada, com linhas ou colunas linearmente dependentes, e não pode ser invertida.

Este problema pode ser contornado ao se analisar todas as $2^N - 1$ possíveis combinações exaustivamente ou então pela seleção de um subconjunto de componentes de acordo com o seguinte método:

- i. Ordenar os componentes em ordem crescente de acordo com seu erro de generalização;
- ii. Selecionar o melhor componente como primeiro componente do *ensemble*;
- iii. Adicionar o próximo membro da sequência ao *ensemble* se a inequação

$$(2N + 1)MSE(\hat{f}_N) > 2 \sum_{i \neq \text{novos}} E[m_{\text{novos}} m_i] + E[m_{\text{novos}}^2] \quad (3.16)$$

for satisfeita, onde N é a quantidade de membros já existente no *ensemble*, \hat{f}_N é a previsão do *ensemble*, $E[.]$ é a esperança matemática e m (*missfit*) é o erro entre o preditor e o conjunto de validação, $m = \hat{f}(x) - f(x)$.

3.9 MÉTODOS DE COMBINAÇÃO DE COMPONENTES DO COMITÊ

Nesta seção são apresentadas as equações e algoritmos utilizados pelos métodos de combinação utilizados.

3.9.1 Média simples (*Basic Ensemble Method* - BEM)

Perrone e Cooper (1993) definem a média simples como método de *ensemble* básico, o qual é dado por

$$f_{BEM}(x) = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (3.17)$$

onde $f_{BEM}(x)$ é a previsão do *ensemble* para o instante x , N é a quantidade de componentes do *ensemble* e $f_i(x)$ é a saída do i – ésimo componente.

3.9.2 Moda

A moda é definida como o valor mais frequente em um conjunto de dados finito (discreto) e sua vantagem em relação à média simples se deve a ser menos sensível a erros extremos, ou seja, se um dos valores do conjunto do qual se pretende calcular a moda diferir muito dos demais o resultado é pouco afetado. Contudo, para variáveis contínuas como as séries temporais, os elementos da série podem assumir infinitos valores, o que torna inviável o cálculo da moda. Isto posto tem-se duas possibilidades, a discretização dos valores da série ou a estimação da densidade das distribuições de probabilidades dos valores do conjunto. Kourentzes *et al.* (2014) utiliza a estimação da densidade de distribuições de probabilidade contínuas, denominadas *kernel*, como o valor da moda para combinação de comitês.

A FIGURA 12 ilustra o procedimento de cálculo da estimação da densidade das distribuições de probabilidade *kernel*. Cada ponto marcado no eixo das abcissas corresponde a saída de uma *wavenet* do comitê em um determinado horizonte de previsão, em torno do qual uma distribuição de probabilidade *kernel* gaussiana é representada pelas linhas tracejadas. Quando as estimativas das distribuições *kernel* de todas as previsões são calculadas, procede-se ao cálculo da densidade, que é a soma de todas as distribuições de probabilidade individuais e é representada pela linha contínua. O valor da moda é a projeção do valor máximo da densidade no eixo das abcissas.

Ainda, de acordo com Kourentzes *et al.* (2014), são detalhados os passos para se estimar a densidade de probabilidade *kernel* incluindo a escolha da função *kernel* e da largura de banda. Neste trabalho é utilizada a função *ksdensity* do ambiente computacional Matlab para este propósito, a qual é baseada no trabalho de Bowman e Azzalini (1997).

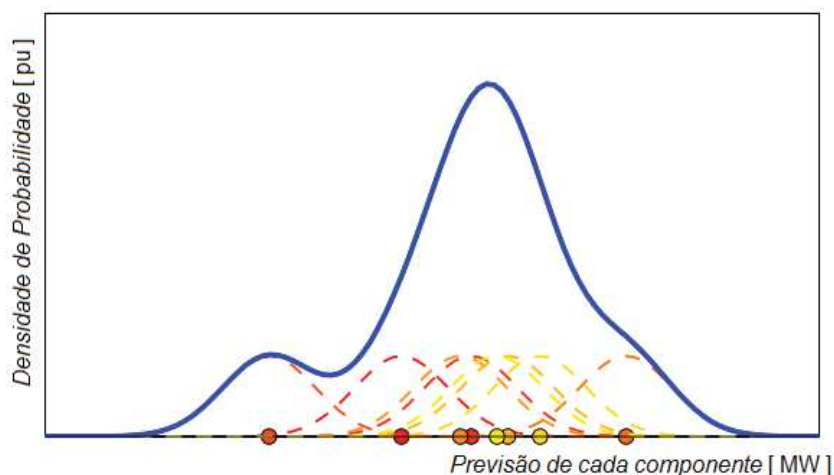


FIGURA 12 – ESTIMAÇÃO DA DENSIDADE KERNEL

FONTE: Adaptado de Kourntzes *et al.* (2014)

3.9.3 Mediana

A mediana aplicada na combinação das previsões é realizada primeiramente através do ordenamento das previsões de cada um dos m modelos para cada horizonte k , e na sequência a mediana é dada pelo valor central no caso em que m é ímpar ou pela média dos dois valores centrais caso m seja par.

Por exemplo, a mediana do vetor [1 2 3 4] é 2,5 enquanto que a mediana do vetor [1 2 3 4 99] é três. Este exemplo expõe a vantagem da mediana sobre a média simples com relação a ser menos sensível a *outliers*, o qual no caso é o valor 99.

3.9.4 Generalização empilhada (*stacked generalization*)

Este método foi proposto por Wolpert (1992) e é ilustrado pela FIGURA 13. Os modelos são separados em dois níveis, no nível zero as entradas de treinamento são aplicadas aos m preditores do *ensemble* e as saídas são utilizadas como entradas no estimador do nível 1, o qual utiliza as saídas do conjunto de dados de treinamento para estimar o modelo combinador do *ensemble*.

Uma vez estimado o modelo combinador as entradas do conjunto de validação são fornecidas aos preditores individuais do *ensemble*, e as saídas geradas pelos preditores individuais são organizadas em m colunas.

Cada coluna com a saída de um preditor individual é fornecida como uma entrada para o modelo combinador, o qual processa os valores fornecidos e provê na sua saída o valor previsto, o qual é a saída do *ensemble*.

3.10 DIFERENÇA ENTRE GENERALIZAÇÃO EMPILHADA, *BAGGING* E *ENSEMBLE*

Esta seção é dedicada a deixar clara a diferença entre os conceitos de generalização empilhada, *bagging* e *ensemble*. A generalização empilhada, ou *stacked generalization*, é um método de combinação de previsões por meio de um modelo criado exclusivamente para esta tarefa. Já o *bagging* é o nome dado ao processo de construção de um *ensemble* através da geração de preditores pelo método do *bootstrapping* e combinados utilizando-se a média simples das previsões que foi proposto por Breiman (1996).

Os comitês de máquinas se dividem basicamente em dois tipos: a mistura de especialistas e o *ensemble*. Basicamente a mistura de especialistas consiste em dividir o objetivo principal em sub tarefas que são delegadas e resolvidas por modelos especialistas, os resultados são então compilados com base no valor das entradas para formar uma solução única ao final, já no *ensemble* cada modelo é gerado para resolver o objetivo principal, e suas soluções são combinadas ao final.

Como um *ensemble* é um dos tipos de comitês de máquinas e neste trabalho não se faz uso do tipo mistura de especialistas, os termos “*ensemble*” e “comitê de máquina” são utilizados indistintamente, tendo o mesmo significado.

Considere por exemplo o problema de previsão de demanda de energia da próxima semana, ao se utilizar a mistura de especialistas divide-se o problema em sete novos problemas, criando um modelo para prever a demanda de cada dia diferente, já ao usar um *ensemble* cada modelo é criado para prever a demanda da semana inteira e suas previsões semanais é que são combinadas ao final.

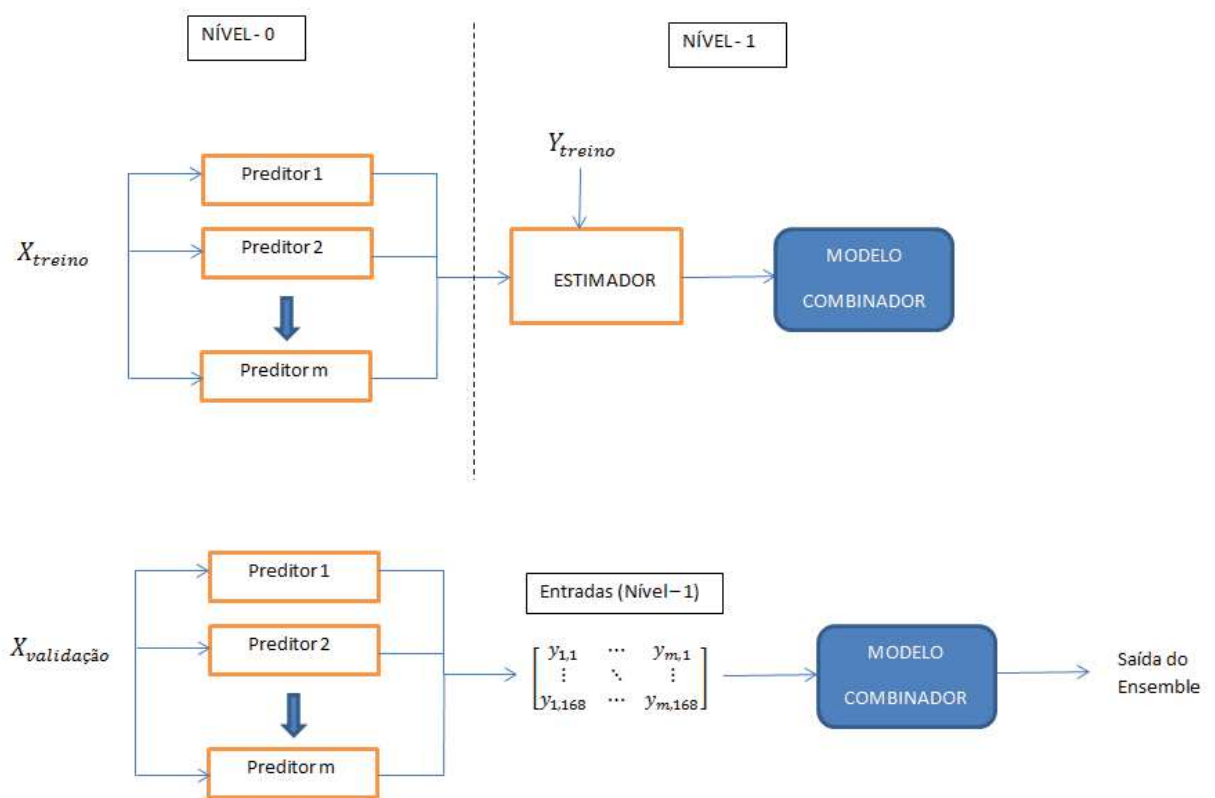


FIGURA 13 – GENERALIZAÇÃO EMPILHADA

FONTE: O autor (2016)

4 RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados inicialmente os resultados do pré-processamento dos conjuntos de dados, na sequência são comparados os resultados de diferentes abordagens para comitês de *wavenets* e, por fim, o comitê com melhor desempenho é comparado com outros algoritmos.

4.1 PRÉ-PROCESSAMENTO DOS DADOS

A série GEFCom possui 8760 valores de demanda horária em MW do ano de 2005 conforme o gráfico na FIGURA 14. Ao analisá-la percebe-se uma frequência baixa sobreposta a frequências altas, provavelmente referentes aos padrões de estações do ano, semanais e diários, além de possíveis ruídos.

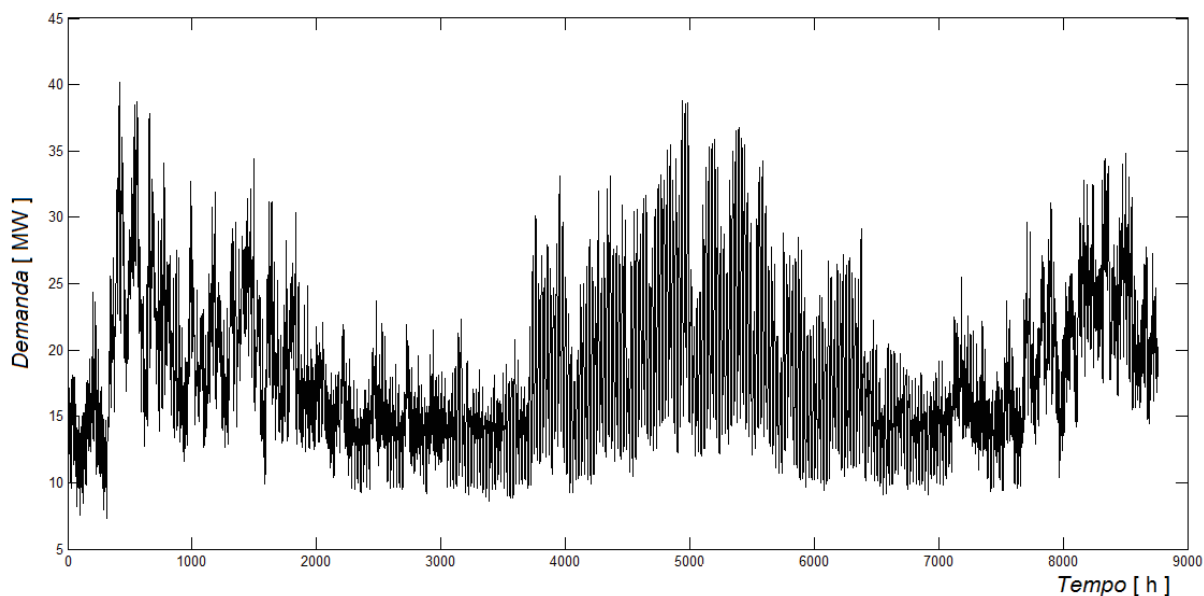


FIGURA 14 - SÉRIE DE DEMANDA HORÁRIA GEFCOM NO ANO DE 2005

FONTE: O autor (2016)

Da mesma forma a série Itália possui 8760 valores de demanda horária em GW conforme o gráfico da FIGURA 15. Nesta série, diferentemente da anterior, é

possível visualizar claramente os padrões mensais, contudo ainda existe uma baixa frequência sobreposta a altas frequências.

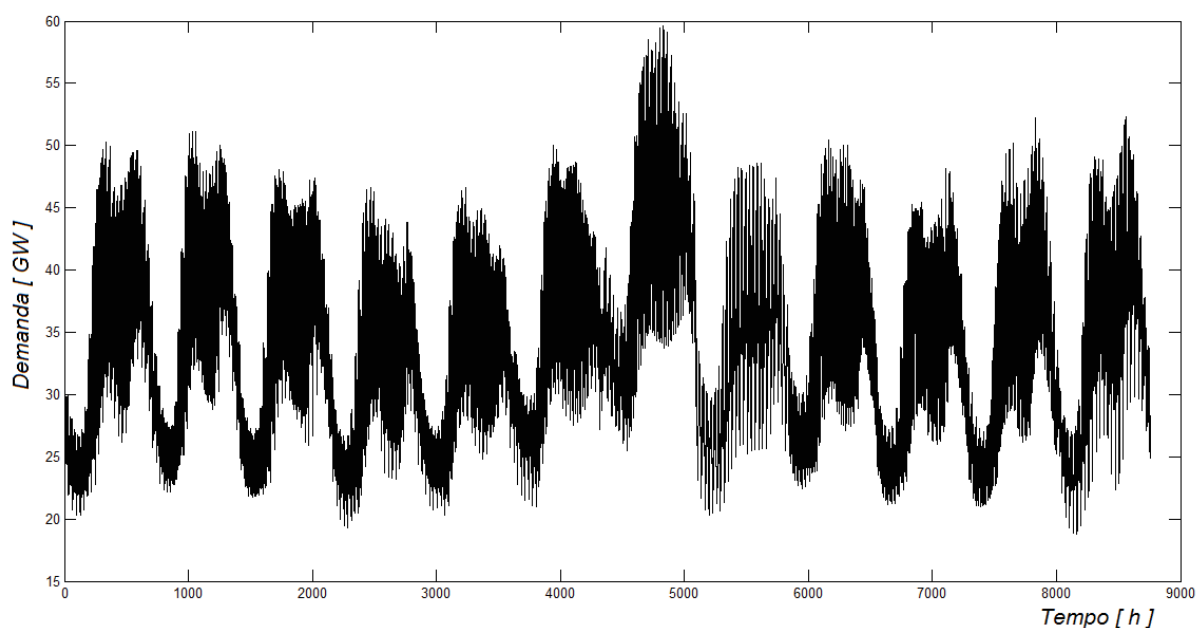


FIGURA 15 - SÉRIE DE DEMANDA HORÁRIA DA ITÁLIA NO ANO DE 2005

FONTE: O autor (2016)

Ao se aplicar o teste de não linearidade proposto por Billings e Voon (1983) sobre as séries GEFCom e Itália, obtêm-se os gráficos da FIGURA 16 e FIGURA 17 respectivamente. Os valores de correlação cruzada utilizados nos índices estão situados fora do intervalo de confiança para as duas séries, indicando claramente que são não lineares.

TABELA 1 - ESTATÍSTICA DA TRANSFORMAÇÃO BOX-COX SOBRE AS SÉRIES

Série	Média	Mediana	Máximo	Mínimo	Desvio Padrão	Assimetria	Curtose
GEFCom Original	18,38	17,11	40,21	7,32	5,77	0,77	3,11
GEFCom Transformada	2,24	2,23	2,70	1,67	0,18	0,01	2,44
Itália Original	35,89	35,25	59,65	18,79	8,41	0,15	2,00
Itália Transformada	8,86	8,85	11,82	6,11	1,19	-0,04	1,94

FONTE: O autor (2016)

Constatada a não linearidade das séries aplica-se a transformação Box-Cox nos dois conjuntos de dados e obtêm-se as estatísticas apresentadas na TABELA 1. O desvio padrão é reduzido significativamente para as duas séries após a

transformação, na série GEFCom inicialmente ele representa aproximadamente 31% da média e 8% após a transformação, no caso da Itália o desvio padrão inicial é de aproximadamente 23% e cai para 13% após a transformação.

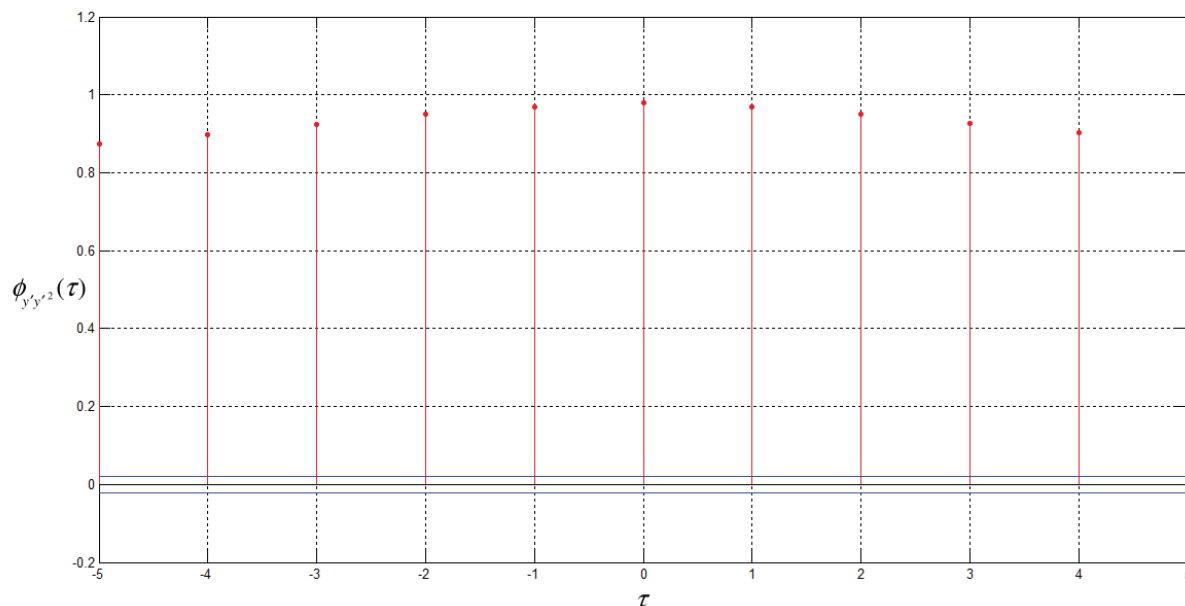


FIGURA 16 - TESTE DE NÃO LINEARIDADE PARA A SÉRIE GEFCOM

FONTE: O autor (2016)

NOTA: As linhas azuis demarcam o intervalo de confiança enquanto as barras vermelhas indicam o valor do teste de correlação cruzada

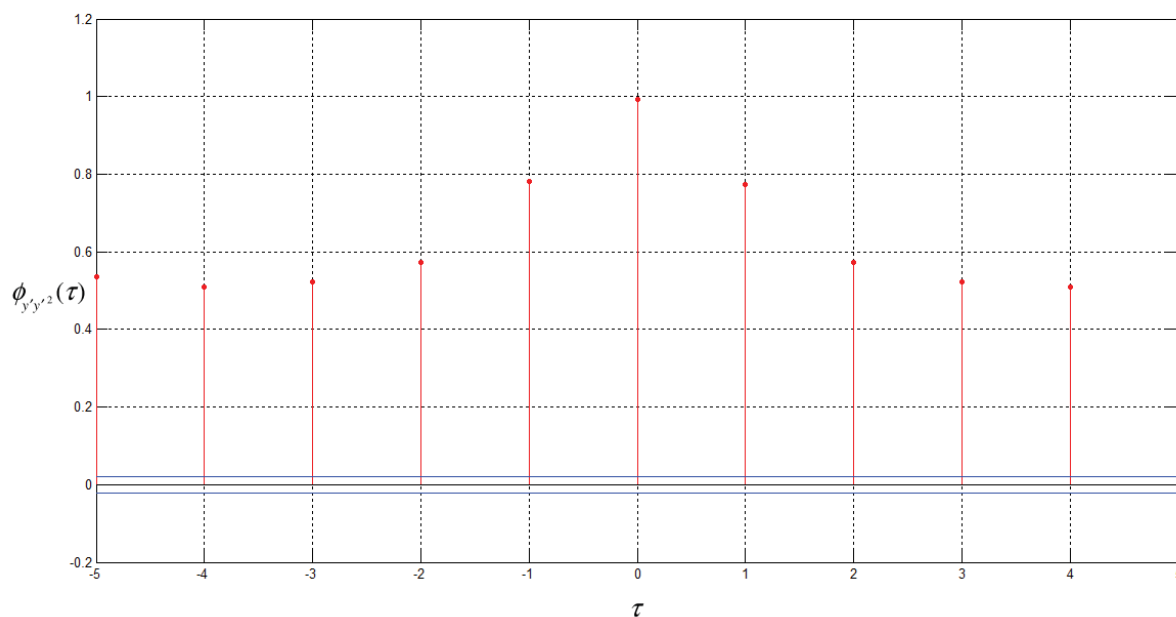


FIGURA 17 - TESTE DE NÃO LINEARIDADE PARA A SÉRIE ITÁLIA

FONTE: O autor (2016)

NOTA: As linhas azuis demarcam o intervalo de confiança enquanto as barras vermelhas indicam o valor do teste de correlação cruzada

Outro benefício da transformação dos dados é a redução da assimetria que pode ser visualizada nos histogramas da FIGURA 18 e FIGURA 19. Além disso, a curtose, apesar de não ter reduzido muito com a transformação, ficou abaixo de três, o que indica que a série é pouco sensível a *outliers*.

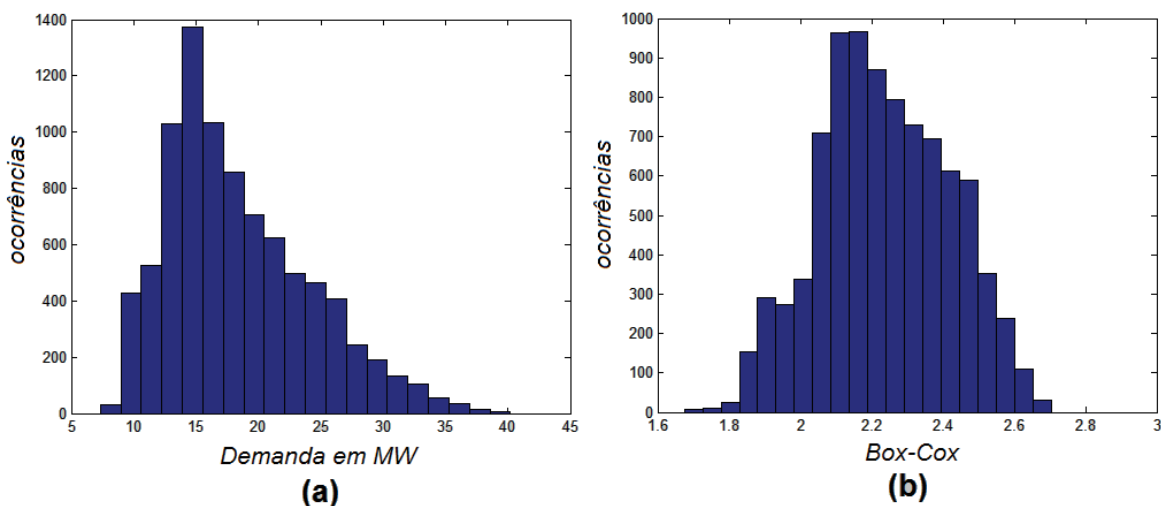


FIGURA 18 - HISTOGRAMA DA SÉRIE GEFCOM (A) ANTES E (B) APÓS A TRANSFORMAÇÃO
 FONTE: O autor (2016)

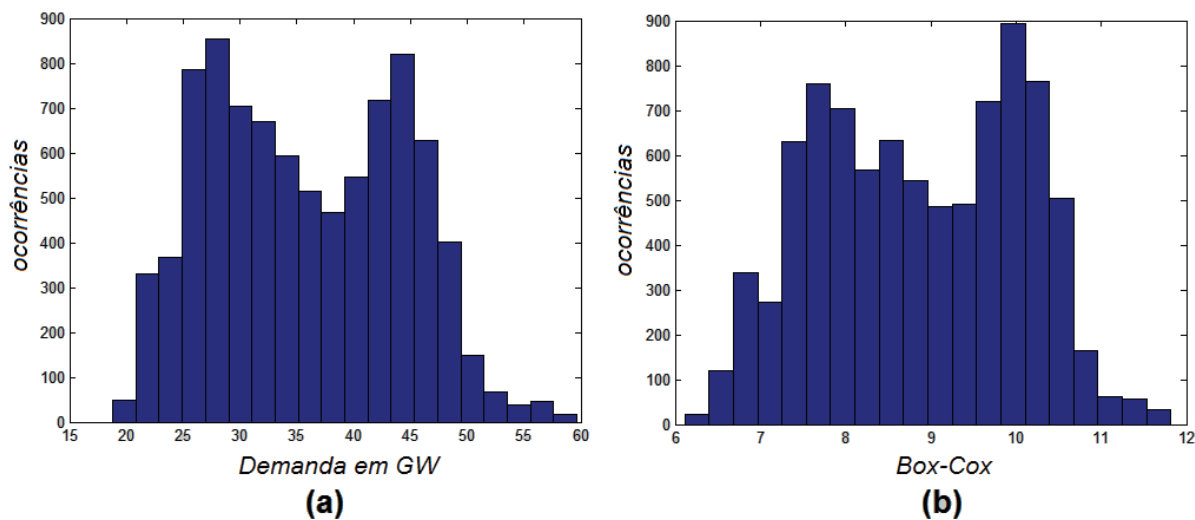


FIGURA 19 - HISTOGRAMA DA SÉRIE ITÁLIA (A) ANTES E (B) APÓS A TRANSFORMAÇÃO
 FONTE: O autor (2016)

Logo após a transformação Box-Cox os dados são normalizados e as séries ficam conforme mostrado na FIGURA 20. A partir dos dados normalizados precisa convertê-los do formato de série para o formato de pares entrada e saída, e para

isso é necessário descobrir o atraso máximo a ser utilizado como variável de entrada para cada série, ou seja, calcular a janela de predição.

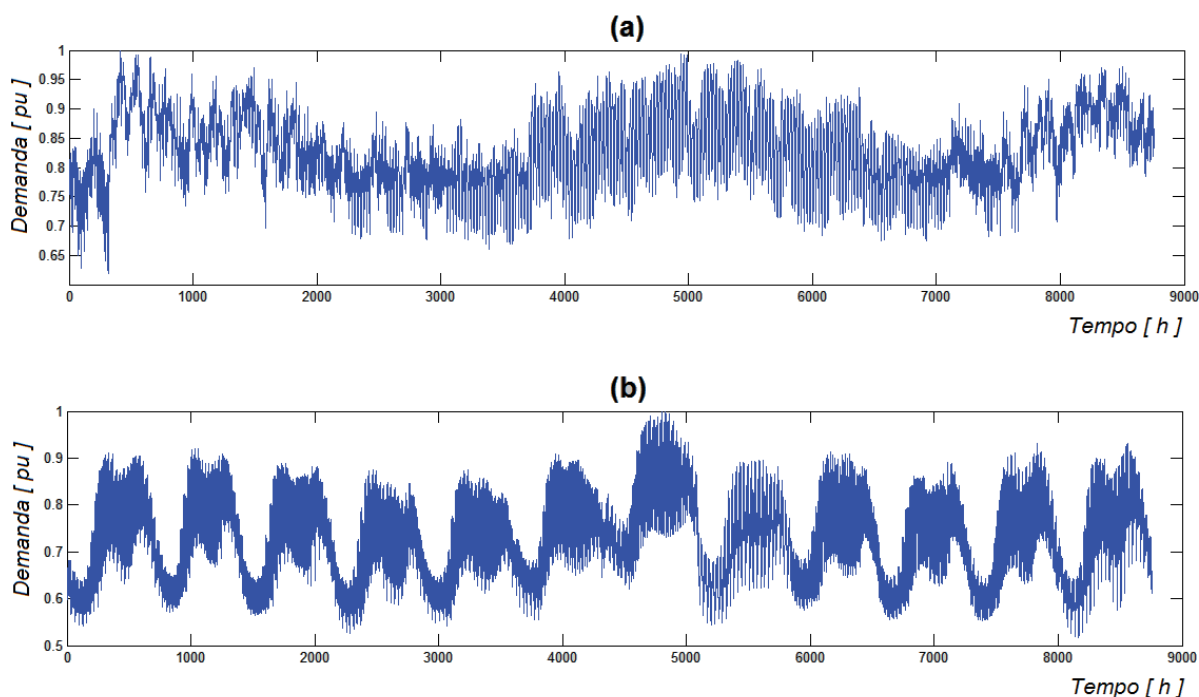


FIGURA 20 - SÉRIES DE DEMANDA HORÁRIA TRANSFORMADAS E NORMALIZADAS (A) GEFCom E (B) ITÁLIA

FONTE: O autor (2016)

Ao se utilizar o método da soma das taxas de redução de erro - SERR proposta por Billings (2013) obtêm-se o gráfico da FIGURA 21 para a série GEFCom e da FIGURA 22 para a série da Itália. O atraso máximo é aquele para o qual não há mais aumento da SERR de uma forma significativa e a curva se torna praticamente horizontal. O atraso máximo determinado para as séries foi de 171 para a série GEFCom e de 69 para a série da Itália. A série GEFCom é arranjada da seguinte forma,

$$\mathbf{X} = \begin{bmatrix} y_1 & y_2 & \cdots & y_{171} \\ y_2 & y_3 & \cdots & y_{172} \\ \vdots & \vdots & \ddots & \vdots \\ y_{8589} & y_{8590} & \cdots & y_{8759} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{172} \\ y_{173} \\ \vdots \\ y_{8760} \end{bmatrix}$$

onde X é a matriz de entradas, na qual cada coluna é uma variável de entrada, e y é um vetor coluna com os valores alvo. Da mesma forma a série Itália fica assim representada,

$$\mathbf{X} = \begin{bmatrix} y_1 & y_2 & \dots & y_{69} \\ y_2 & y_3 & \dots & y_{70} \\ \vdots & \vdots & \ddots & \vdots \\ y_{8691} & y_{8692} & \dots & y_{8759} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_{70} \\ y_{71} \\ \vdots \\ y_{8760} \end{bmatrix}$$

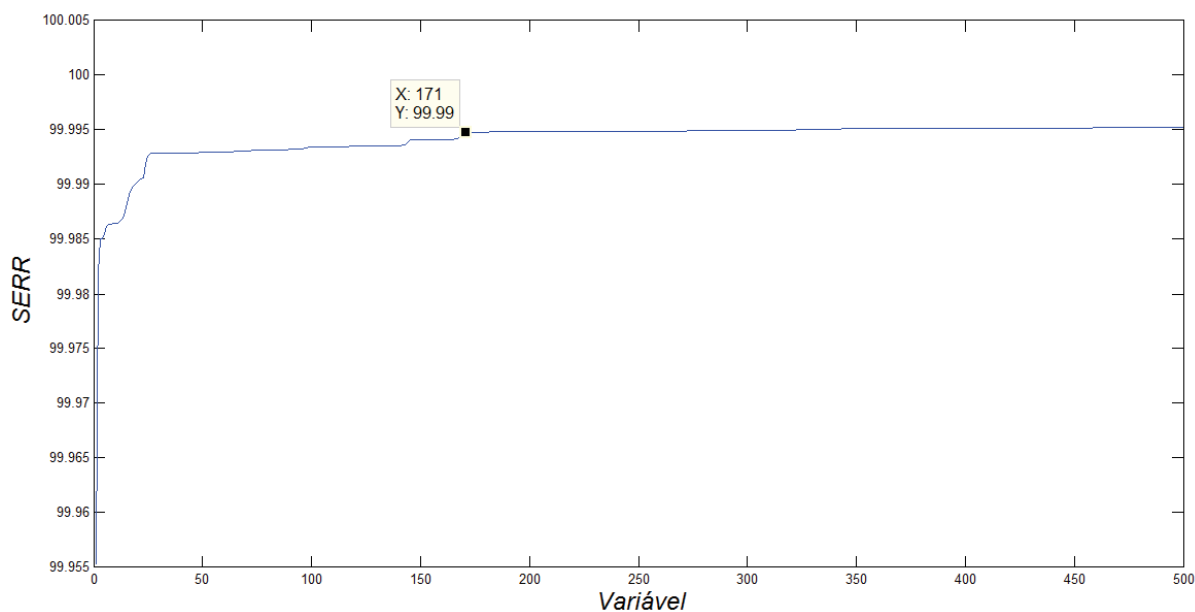


FIGURA 21 - SERR PARA A SÉRIE GEFCOM

FONTE: O autor (2016)

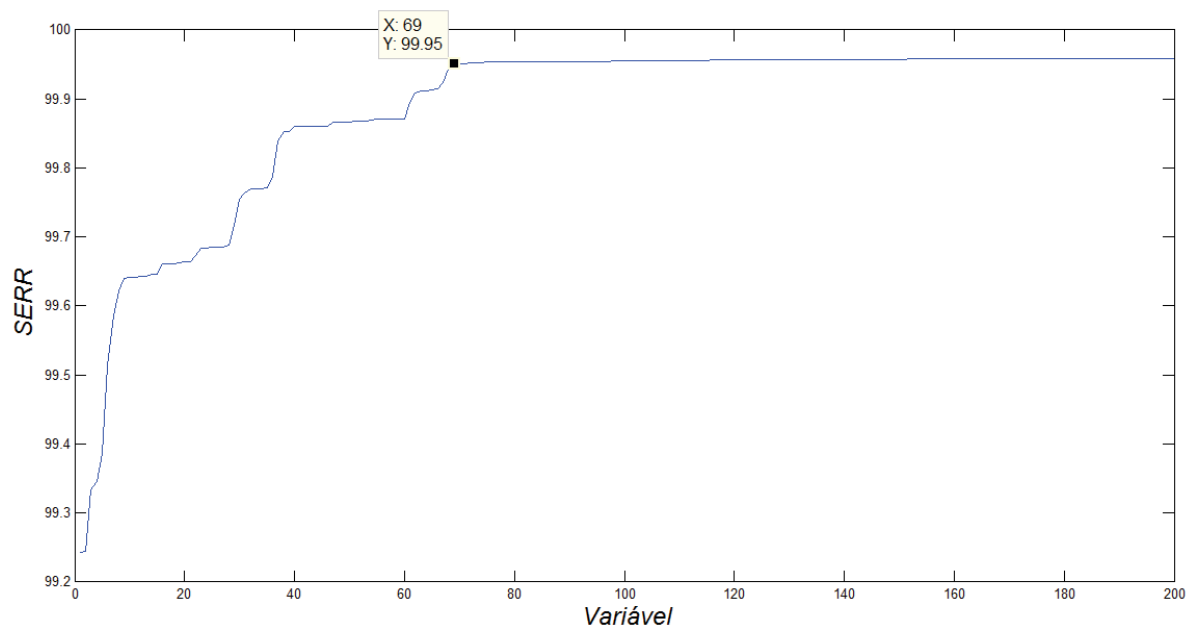


FIGURA 22 - SERR PARA A SÉRIE ITÁLIA

FONTE: O autor (2016)

É possível notar que devido ao atraso máximo ser diferente para cada série são obtidas 8589 amostras para a série GEFCom e 8691 para a série Itália, o que

não é um problema devido a quantidade total ser da ordem de milhares de amostras, contudo caso o espaço amostral seja reduzido passa a existir um compromisso entre a janela de predição e a quantidade de exemplos.

A partir dos dados transformados e normalizados, o próximo passo é realizar a seleção de variáveis para o modelo, ou seja, escolher dentre as colunas da matriz X apenas aquelas essenciais para representar a dinâmica da série. Utiliza-se o algoritmo de seleção sequencial para frente, ou *sequential forward selection* (SFS) e obtêm-se as variáveis mostradas na TABELA 2 para cada série.

TABELA 2 - VARIÁVEIS SELECIONADAS PARA CADA SÉRIE

Série	Atrasos em horas
GEFCom	1, 2, 3, 21, 27
Itália	2, 7, 9, 19, 32, 33, 39, 40, 63

FONTE: O autor (2016)

É possível notar que a série Itália demandou mais variáveis do que a série GEFCom, nove e cinco respectivamente, sendo que é possível representar os modelos preditores tal que

$$y_{GEFCom} = f(y(t-1), y(t-2), y(t-3), y(t-21), y(t-27)) + e(t) \quad (4.1)$$

$$y_{Itália} = f(y(t-2), y(t-7), y(t-9), y(t-19), y(t-32), y(t-33), y(t-39), y(t-40), y(t-63)) + e(t) \quad (4.2)$$

4.2 COMITÊS DE WAVENETS

Nesta seção são apresentados e comparados os resultados dos comitês de *wavenets* utilizando-se os seguintes algoritmos de aprendizado de comitê:

- a) *bagging*;
- b) validação cruzada para partição dos dados;
- c) dizimação de entradas;
- d) seleção construtiva;

- e) combinação pela moda;
- f) combinação pela mediana;
- g) generalização empilhada.

Os experimentos são realizados com o processo de validação cruzada de 10 repetições, ou seja, as amostras são indexadas de 1 a 10 e na primeira iteração as amostras com índice diferentes de “1” (90%) são usadas para o treinamento e aquelas com índice “1” são usadas para teste, e assim sucessivamente até o índice de número 10.

Ao final de cada repetição da validação cruzada é obtido um vetor com 168 valores, que são a média do índice R^2 para cada horizonte de previsão do conjunto de teste, ao término das 10 repetições é calculado o R^2 médio, também com 168 valores, que é a média dos 10 valores obtidos nas repetições para cada horizonte. Para cada comitê é definido como parâmetro a geração de 30 *wavenets*.

As *wavenets* são estimadas pelo ambiente computacional Matlab e está fora do escopo deste trabalho a otimização da estrutura algoritmos de aprendizado para cada conjunto de treinamento. Para o *bagging* é utilizada uma taxa de 100% de reamostragem, ou seja, para cada *wavenet* gerada todo o conjunto de treinamento é reamostrado.

Já para o algoritmo que utiliza a validação cruzada para aprendizado do comitê o conjunto de dados de treinamento é particionado aleatoriamente em 30 subconjuntos independentes, um para cada *wavenet*. A combinação das previsões é realizada pela média simples.

No algoritmo de dizimação de entradas, antes de se gerar cada *wavenet* seleciona-se aleatoriamente uma variável a ser dizimada do conjunto de treinamento, a combinação também é realizada com a média simples.

4.2.1 Previsão para a série GEFCom

O gráfico da FIGURA 23 apresenta os índices de desempenho para o comitê bem como o obtido ao se utilizar uma única *wavenet* como preditor para a série GEFCom. Ao se analisar o gráfico percebe-se que o desempenho dos preditores

decrece à medida que o horizonte de previsão aumenta, também é possível perceber que até a oitava hora aproximadamente os desempenhos são bem próximos, mas a partir daí até o início do quinto dia o comitê que utiliza o método da validação cruzada na partição dos dados para gerar as *wavenets* demonstra superioridade com relação aos demais.

O método da dizimação de entradas obtém o segundo melhor desempenho estando mais próximo do método da validação cruzada do que os demais. O comitê criado com o *bagging* foi melhor do que apenas uma *wavenet* como modelo preditor. A partir do quinto dia, aproximadamente na centésima hora à frente, os comitês não são mais capazes de explicar a dinâmica da série.

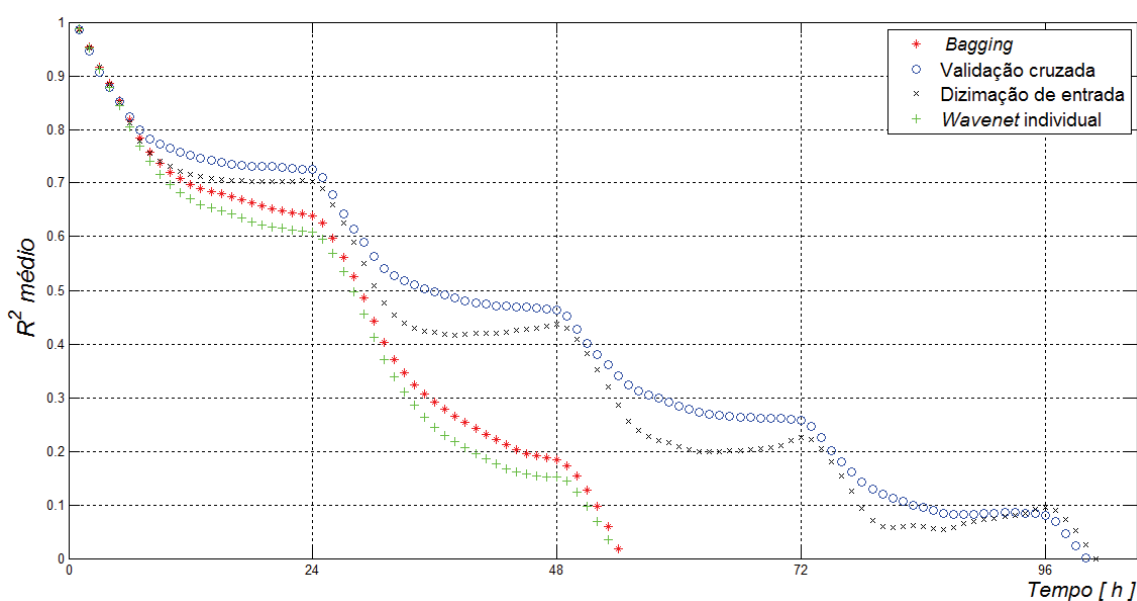


FIGURA 23 - ÍNDICE R^2 MÉDIO DOS COMITÊS PARA A SÉRIE GEFCOM

FONTE: O autor (2016)

Na sequência aplica-se o método de seleção construtiva de variáveis proposto por Perrone e Cooper (1993), sendo que para o processo de seleção são separadas 10% das amostras do conjunto de treinamento para validação. O gráfico da FIGURA 24 compara o comitê de *wavenets* criado com a validação cruzada na geração de componentes, sem e com a etapa de seleção.

Ao contrário do que se espera o desempenho do comitê piora com a etapa de seleção. É possível que este fato se deva ao índice de desempenho considerado na proposição do método de seleção ser diferente do avaliado neste experimento. A

inequação é deduzida analiticamente a partir do erro médio quadrático para decidir se o desempenho do comitê melhora ou piora com a adição do novo componente, contudo o coeficiente de correlação múltipla não considera o erro médio quadrático explicitamente, mas sim uma relação entre o quadrado dos resíduos e a variância dos dados medidos.

A seguir são comparados os métodos de combinação das previsões com a média simples, a moda, a mediana e a generalização empilhada conforme mostrado no gráfico da FIGURA 25. Nota-se no gráfico que a mediana e a moda melhoram o desempenho da previsão, por outro lado a generalização empilhada o degrada. Conclui-se que o melhor comitê de *wavenets* para a série GEFCom é o que gera componentes com o método da validação cruzada, sem seleção e combinação através da moda.

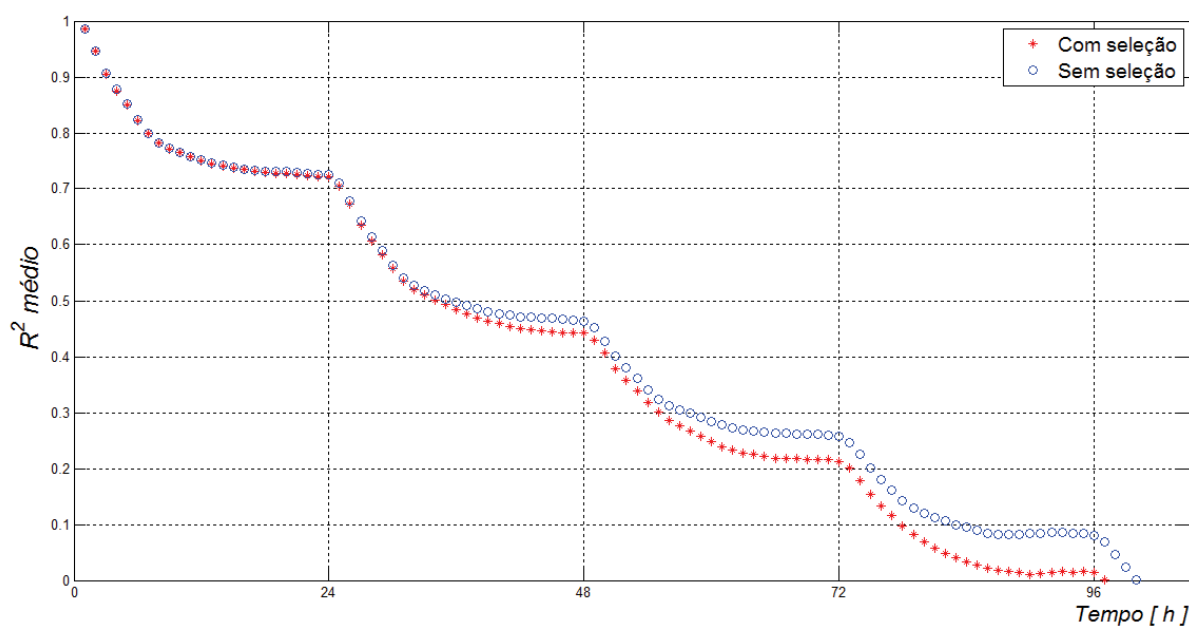


FIGURA 24 - ETAPA DE SELEÇÃO PARA A SÉRIE GEFCOM

FONTE: O autor (2016)

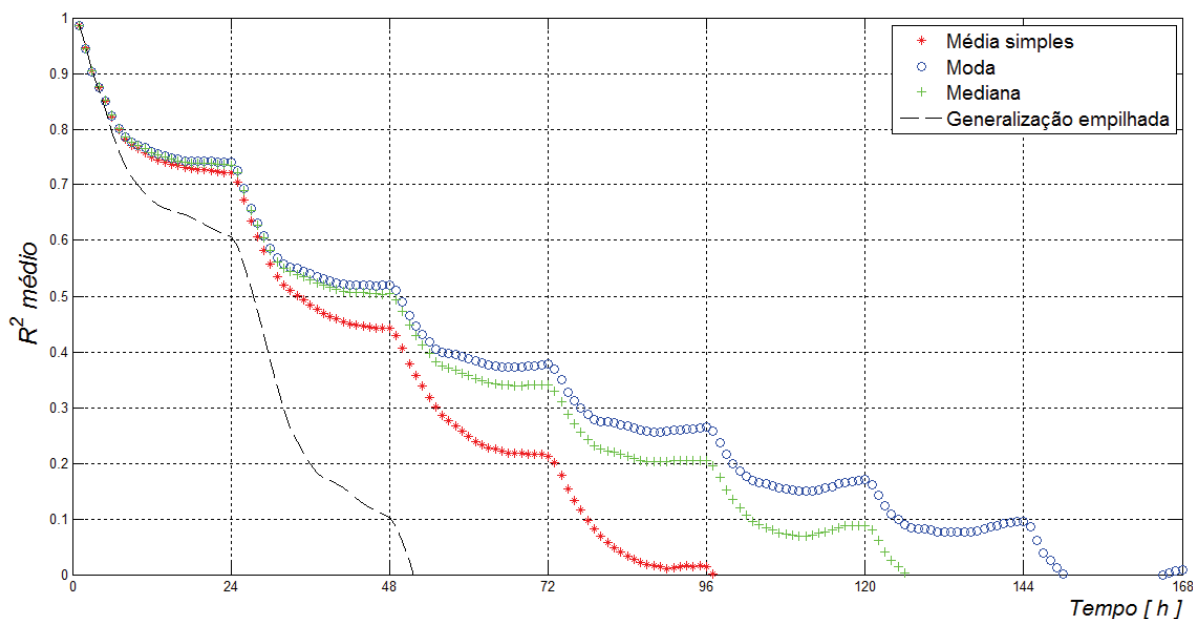


FIGURA 25 - COMPARAÇÃO ENTRE OS MÉTODOS DE COMBINAÇÃO PARA A SÉRIE GEFCom
 FONTE: O autor (2016)

4.2.2 Previsão para a série Itália

O gráfico da FIGURA 26 apresenta os índices de desempenho para o comitê bem como o obtido ao se utilizar uma única *wavenet* como preditor para a série Itália. Ao se analisar o gráfico percebe-se que o desempenho dos preditores decresce à medida que o horizonte de previsão aumenta, também é possível perceber que até a oitava hora aproximadamente os desempenhos são bem próximos, mas a partir daí até o início do quarto dia o comitê que utiliza o *bagging* e o preditor de uma única *wavenet* demonstram superioridade com relação aos demais, já a partir do quarto dia o comitê criado pelo *bagging* apresenta melhor desempenho.

Contrariamente ao resultado para a série GEFCom o comitê que utiliza a validação cruzada para a geração de *wavenets* e a dizimação de entradas obtém o pior desempenho, contudo este resultado pode ser compreendido ao se relembrar do teorema *no free lunch* (Wolpert e MacReady, 1997).

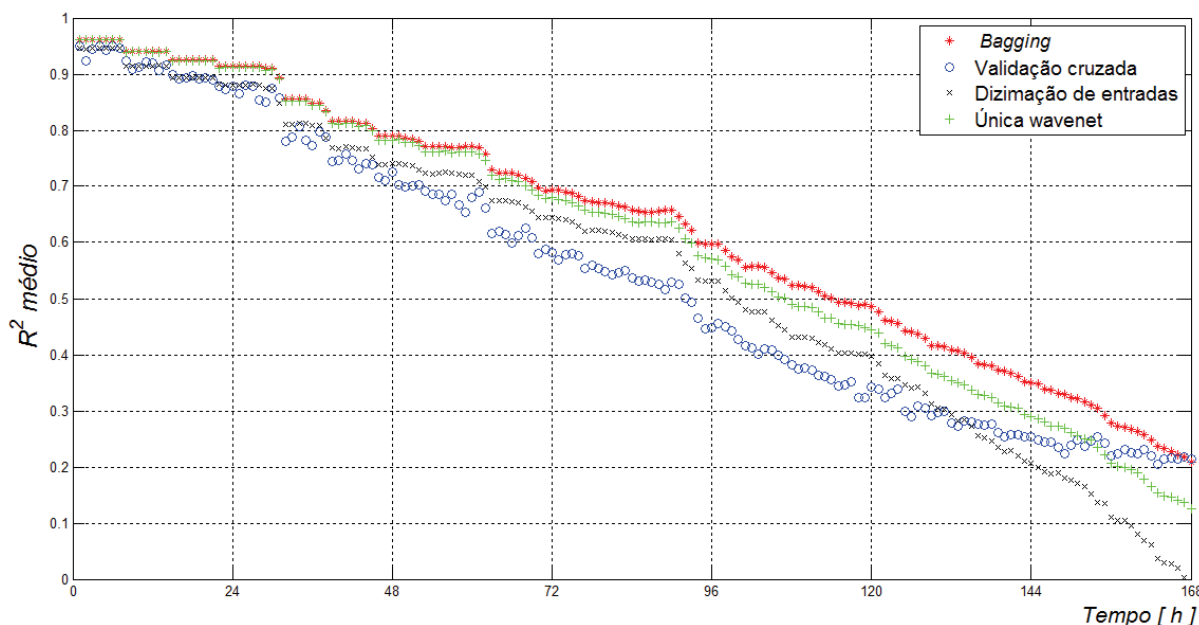


FIGURA 26 - ÍNDICE R^2 MÉDIO DOS COMITÊS PARA A SÉRIE ITÁLIA

FONTE: O autor (2016)

Na sequência aplica-se o método de seleção construtiva de variáveis proposto por Perrone e Cooper (1993) sobre os modelos gerados através de reamostragem com reposição, que foi o que deu origem ao comitê de *wavenets* com melhor desempenho, sendo que para o processo de seleção são separadas 10% das amostras do conjunto de treinamento para validação. O gráfico da FIGURA 27 compara o comitê de *wavenets* criado com a validação cruzada na geração de componentes, sem e com a etapa de seleção. Percebe-se que o desempenho do comitê melhora a partir da metade do 6º dia ao se incluir a etapa de seleção.

A combinação das previsões é realizada com a média simples, a moda, a mediana e a generalização empilhada conforme mostrado no gráfico da FIGURA 28. Nota-se no gráfico que a média simples obtém o melhor desempenho da previsão, e a generalização empilhada o pior. Conclui-se que o melhor comitê de *wavenets* para a série Itália é o que gera componentes com o método da reamostragem por reposição (*bootstrapping*), com seleção e combinação através da média simples.

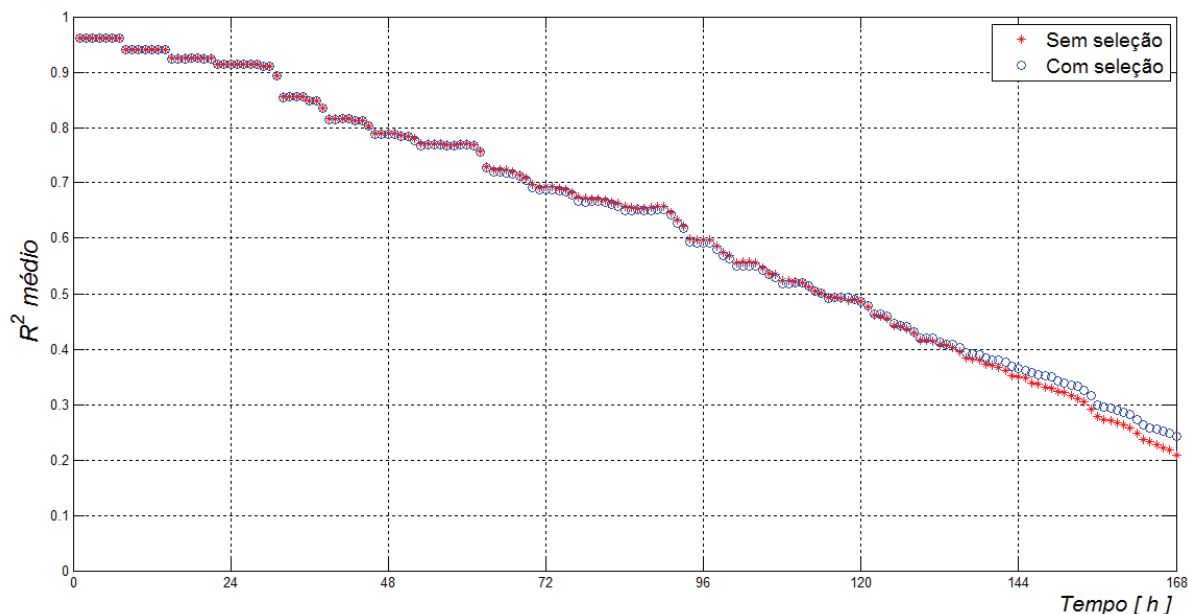


FIGURA 27 - PREVISÃO PARA A SÉRIE ITÁLIA COM E SEM A ETAPA DE SELEÇÃO

FONTE: O autor (2016)

4.3 COMPARAÇÃO COM OUTROS ALGORITMOS

Nesta seção compara-se o melhor comitê de *wavenets* para cada série, identificados na seção anterior, com uma rede neural perceptron multicamadas com função de ativação sigmoide na camada oculta, com a média e com o modelo inocente. O modelo inocente considera que a demanda na próxima semana será idêntica a da semana atual.

A FIGURA 29 exhibe o gráfico do desempenho dos algoritmos comparados para previsão da série GEFCom, e nota-se que até oito horas à frente o comitê de *wavenets* e a rede neural artificial sigmoide (i.e. perceptron multicamadas com função de ativação sigmoide na camada oculta) possuem desempenho praticamente iguais enquanto que o modelo inocente e a média possuem desempenho bem menor. A partir da 8^a hora até a 51^a hora o comitê de *wavenets* possui o melhor desempenho, sendo que a partir daí até a 168^a hora o modelo inocente é superior aos demais.

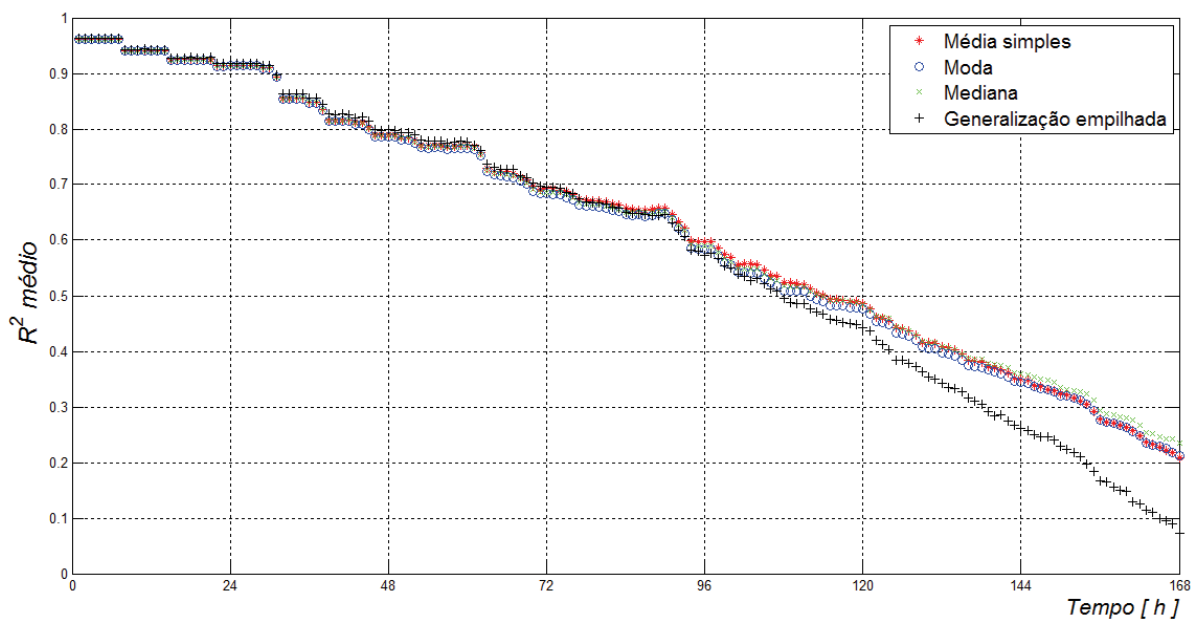


FIGURA 28 - COMITÊS DE WAVENETS COM DIFERENTES MÉTODOS DE COMBINAÇÃO PARA PREVISÃO DA SÉRIE ITÁLIA

FONTE: O autor (2016)

Já a previsão dos modelos comparados para a série Itália é apresentada na FIGURA 30. Diferentemente da série GEFCom, o comitê de *wavenets* e a rede neural sigmoide (i.e. perceptron multicamadas com função de ativação sigmoide na camada oculta) superam a média e o modelo inocente por todo o horizonte de previsão, e possuem desempenhos praticamente iguais até a metade do segundo dia à frente, quando o comitê possui desempenho superior até a 168ª hora à frente. É possível notar do gráfico que o modelo inocente não aparece no gráfico, isto é devido a possuir somente valores negativos do índice R^2 .

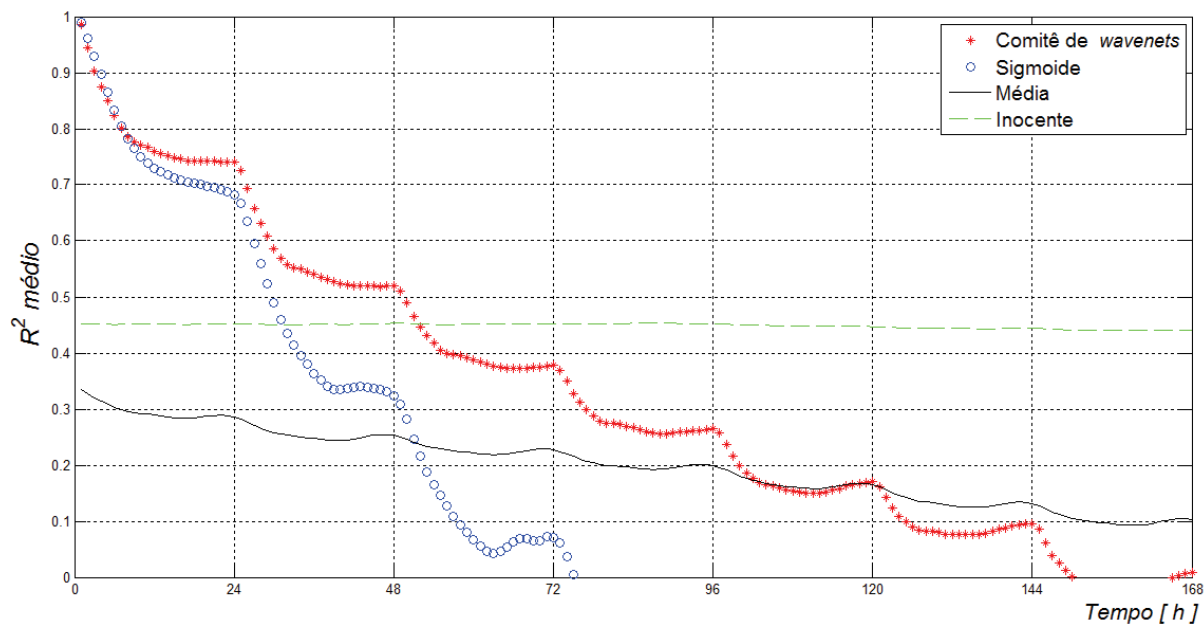


FIGURA 29 - COMPARAÇÃO DO COMITÊ DE WAVENETS COM OUTROS ALGORITMOS PARA PREVISÃO DA SÉRIE GEFCOM

FONTE: O autor (2016)

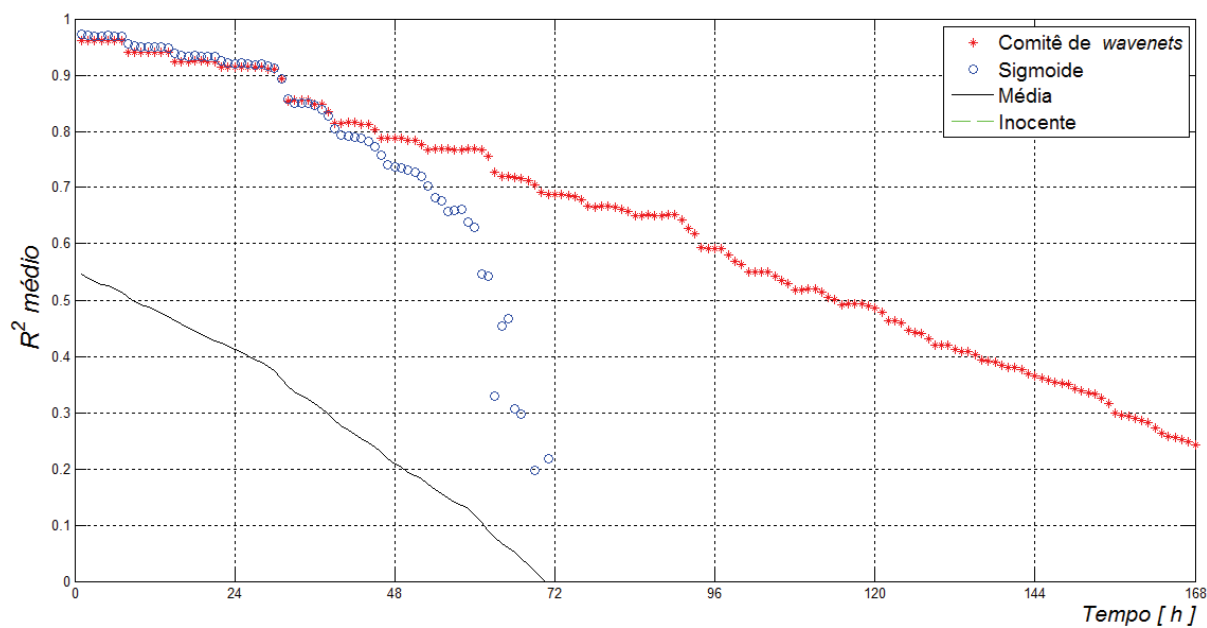


FIGURA 30 - COMPARAÇÃO DO COMITÊ DE WAVENETS COM OUTROS ALGORITMOS PARA PREVISÃO DA SÉRIE ITÁLIA

FONTE: O autor (2016)

4.4 TESTES DE CORRELAÇÃO

A FIGURA 31 e a FIGURA 32 apresentam o resultado dos testes de correlação para os resíduos das previsões dos melhores comitês de *wavenets* para a série GEFCom e Itália respectivamente. Como os modelos são avaliados com um processo de 10 validações cruzadas, e cada validação possui várias amostras, os resíduos são tomados como a média dos resíduos de todas as amostras. Outra característica aplicável a séries temporais é que não é possível analisar a correlação entre os resíduos e as entradas, pois as mesmas são valores passados da saída e, conseqüentemente, dos próprios resíduos.

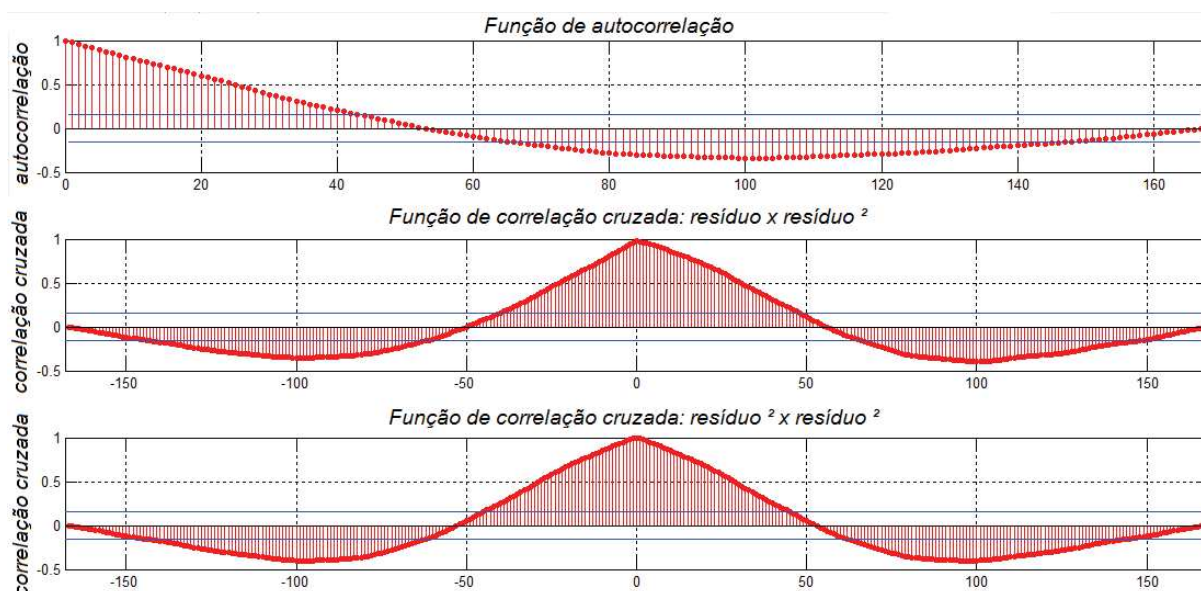


FIGURA 31 - TESTES DE CORRELAÇÃO PARA OS RESÍDUOS DE PREVISÃO DA SÉRIE GEFCOM

FONTE: O autor (2016)

O objetivo desta análise é identificar se há correlação entre os valores dos resíduos da série de forma a identificar termos que deveriam estar no modelo. É possível interpretar nos gráficos que há coeficientes de correlação fora do intervalo de confiança de 95%, representado pelas linhas azuis.

Quando os resíduos não possuem correlação é esperado que a autocorrelação entre os resíduos e a autocorrelação entre os quadrados dos resíduos seja diferente de zero para qualquer atraso, contudo espera-se que a

correlação cruzada entre os resíduos e o quadrado dos resíduos seja nula para qualquer atraso, ou seja, esteja no intervalo de confiança, o que não é verificado na FIGURA 31 e FIGURA 32.

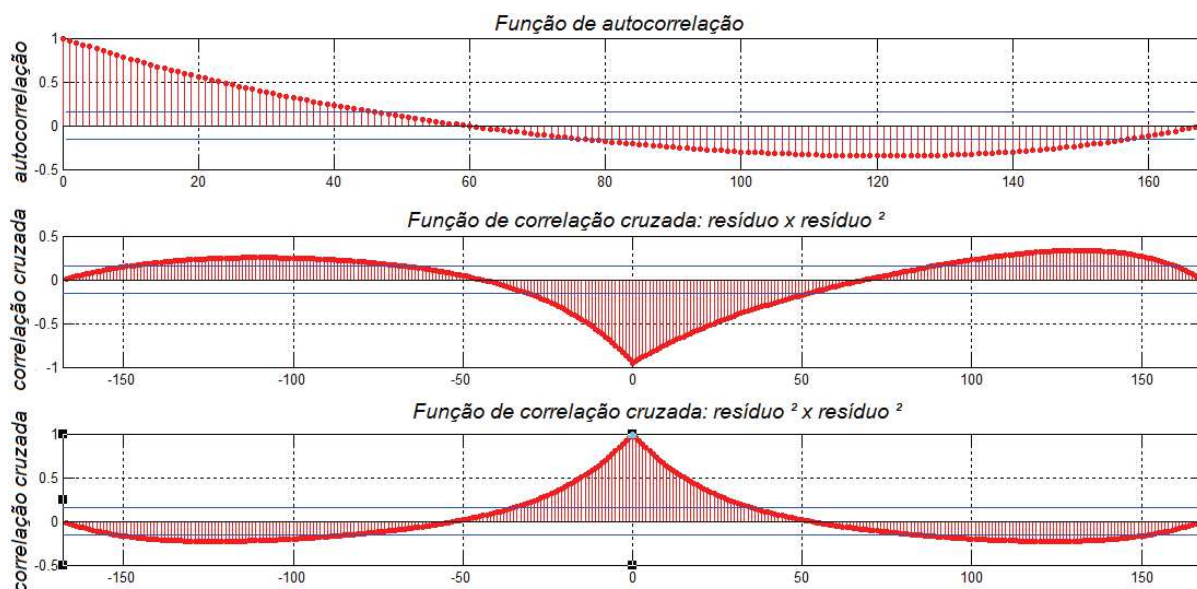


FIGURA 32 - TESTES DE CORRELAÇÃO PARA OS RESÍDUOS DE PREVISÃO DA SÉRIE ITÁLIA
 FONTE: O autor (2016)

Conclui-se, portanto, que apesar dos comitês de *wavenet* superarem o desempenho dos demais modelos comparados, ainda deve ser complementado com um modelo que represente os resíduos. Apenas para ilustração das previsões escolhe-se uma amostra aleatória de cada série, conforme pode ser visto na FIGURA 33 e FIGURA 34.

Nota-se na FIGURA 33 que a previsão é razoável até a oitava hora à frente quando então os erros se acumulam rapidamente e a previsão passa a divergir muito dos dados reais, já na FIGURA 34 a previsão é razoável até a 48ª hora à frente quando então passa a divergir. Ao se consultar a FIGURA 29 para a série GEFCom e a FIGURA 30 para a série Itália, percebe-se que o valor de R^2 é aproximadamente 0,8 para este limiar, ou seja, previsões abaixo deste limiar não aparentam ser razoáveis para serem utilizadas em qualquer aplicação, contudo, ainda assim apresentam melhor precisão do que os outros modelos considerados.

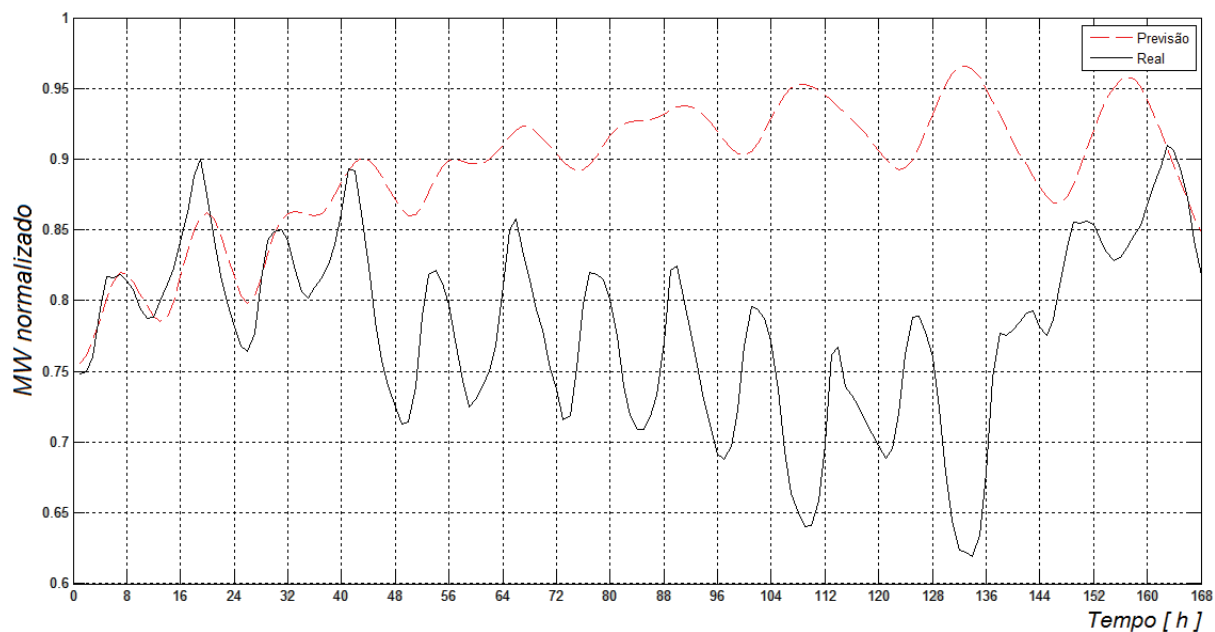


FIGURA 33 - AMOSTRA DE PREVISÃO PARA A SÉRIE GEFCOM

FONTE: O autor (2016)

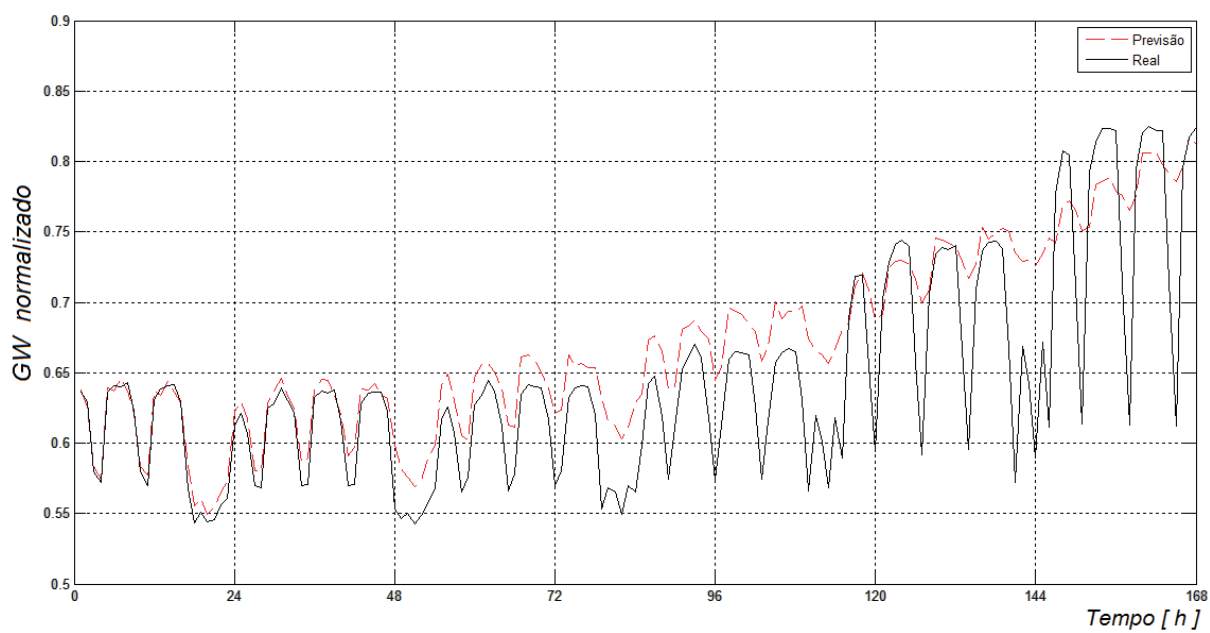


FIGURA 34 - AMOSTRA DE PREVISÃO PARA A SÉRIE ITÁLIA

FONTE: O autor (2016)

5 CONCLUSÕES

Foi realizada uma revisão da literatura específica de previsão de demanda, das *wavenets*, das máquinas de comitê e do pré-processamento dos dados para identificação. Os principais algoritmos identificados foram agrupados cronologicamente em categorias que consideram sua aplicação, abordagem ou similaridade.

Obteve-se uma sistematização da experimentação ao se estabelecer uma sequência de pré-processamento dos dados, caracterização do problema, algoritmos de aprendizado de modelos, de aprendizado de comitê, estratégias de previsão e de validação de modelos que permitem a reprodução e extrapolação do trabalho para outros problemas.

Inicialmente os dados foram tratados para remoção da tendência e normalização através da transformação Box-Cox, e foi possível perceber a redução da assimetria e do desvio padrão proporcionalmente à média dos valores. Além disso as séries foram transformadas em conjuntos de dados de entrada e saída para criação dos exemplos utilizados no aprendizado supervisionado e nos testes de generalização.

Este trabalho abordou a aplicação de comitês de *wavenets* para a previsão de demanda de energia elétrica de curto prazo de duas séries temporais. Mais especificamente foram implementados os algoritmos de reamostragem com reposição, validação cruzada e dizimação de entradas para a geração de *wavenets* para o comitê, o método de seleção construtiva para a etapa de seleção e a média simples, moda, mediana e generalização empilhada na etapa de combinação das previsões. Ainda, as previsões dos comitês de *wavenets* foram comparadas com as previsões de uma *wavenet* simples, rede neural artificial sigmoide (i.e. perceptron multicamadas com função de ativação sigmoide), com a média da última semana e com o modelo inocente.

Os modelos obtidos têm as previsões avaliadas pelo coeficiente de correlação múltipla, R^2 , e por testes de correlação entre os resíduos dos testes decorrentes de um processo de validação cruzada com 10 divisões do conjunto de

dados da série. É observada uma tendência à diminuição de R^2 com o aumento do horizonte de previsão. Os resultados obtidos indicam que é possível obter previsões bem precisas de acordo com o horizonte de previsão e algoritmos de aprendizado de comitês considerados.

Os resultados obtidos para previsão da série GEFCom indicam que o melhor comitê de *wavenets* é aquele obtido pela geração de modelos pela validação cruzada, sem seleção e combinação pela moda. A previsão do comitê de *wavenets* obtida deste conjunto de métodos supera os demais algoritmos comparados até o final do segundo dia à frente, quando o modelo inocente passa a ser superior aos demais.

Já os resultados obtidos para previsão da série Itália indicam que o melhor comitê de *wavenets* é aquele obtido pela geração de modelos por meio da reamostragem com reposição (*bootstrapping*), com seleção e combinação pela média simples. A previsão do comitê de *wavenets* obtido deste conjunto de métodos supera os demais algoritmos durante todo o horizonte de previsão.

Os testes de correlação dos resíduos não foram satisfeitos, para nenhuma das séries, e por isso é possível concluir que existem termos que devem ser incluídos em um modelo do erro para complementar os comitês de *wavenets*. Das amostras de teste de previsão também é possível concluir que as previsões para a série GEFCom são razoáveis até a 8ª hora à frente, quando o acúmulo dos erros de previsão se torna muito grande, já para a série Itália as previsões são razoáveis até 48 horas à frente, constatou-se que o limiar de razoabilidade das previsões equivale a um R^2 mínimo de 0,8.

Os experimentos deste trabalho consideraram duas séries anuais de demanda, consideraram métodos convencionais para aprendizado de comitê e também selecionou alguns modelos para comparação. Em linhas gerais pode-se dizer que um comitê de *wavenets* supera o desempenho dos demais modelos comparados, o que não significa que serão melhores para todos os problemas ou que são melhores do que qualquer outro modelo.

Esta vantagem não é destacada no curtíssimo prazo, onde a rede neural artificial sigmoide (i.e. perceptron multicamadas com função de ativação sigmoide na

camada oculta) é um pouco superior, mas sim por ter um desempenho similar no curtíssimo prazo e se destacar à medida que o horizonte de previsão aumenta.

Para trabalhos futuros sugere-se ampliar estas escolhas para outras séries de demanda de energia, utilizar diferentes abordagens para aprendizado de comitê e adicionar outros modelos para comparação.

Além das sugestões anteriores também é possível avaliar o impacto de diferentes métodos para seleção de variáveis, como a análise de componentes principais, análise fatorial, análise de componentes independentes e também metaheurísticas aplicadas a esta tarefa.

As *wavenets* podem ser estimadas por algoritmos de aprendizado baseado em metaheurísticas (e.g. algoritmos evolutivos), podem ter a quantidade de *wavelons* otimizada e até mesmo possuem outro tipo de *wavelet* nos *wavelons*, como a do tipo Morlet ou Haar. Além disso, existem os comitês heterogêneos, nos quais é possível combinar *wavenets* com outras classes de modelos (e.g. *wavenets* e perceptron multicamadas).

A estratégia de previsão adotada foi a iterativa, contudo podem ser testadas a estratégia direta, que gera um modelo para previsão de cada horizonte pretendido e a que utiliza um modelo de múltiplas entradas e múltiplas saídas, no qual cada saída é o valor em um horizonte de previsão. Os primeiros resultados deste trabalho foram publicados por Ribeiro *et al.* (2016).

REFERÊNCIAS

ADHIKARI, R.; VERMA, G.; KHANDELWAL, I. **A model ranking based selective ensemble approach for time series forecasting**. *Procedia Computer Science*, v. 48, pp. 14-21, 2015.

AGUIRRE, L. A.; RODRIGUES D. D.; LIMA S. T.; MARTINEZ C. B. **Dynamical prediction and pattern mapping in short-term load forecasting**. *International Journal of Electrical Power & Energy Systems*, v. 30, n. 1, p. 73-82, 2008.

AHMAD, Z.; ZHANG, J. **Selective combination of multiple neural networks for improving model prediction in nonlinear systems modeling through forward selection and backward elimination**. *Neurocomputing*, v. 72, n. 4, p. 1198-1204, 2009.

AKSELA, M. Comparison of classifier selection methods for improving committee performance. *Proceedings of the 4th International Workshop on Multiple Classifier Systems*, Guildford, Reino Unido, pp. 84-93, 2003.

ALEXANDRIDIS, A. K.; ZAPRANIS, A. D. **Wavelet neural networks: a practical guide**. *Neural Networks*, v. 42, pp. 1-27, 2013.

ALEXANDRIDIS, A. K.; ZAPRANIS, A. D. **Wavelet neural networks: with applications in financial engineering, chaos, and classification**. John Wiley & Sons, 2014.

ANDALIB, A.; ATRY, F. **Multi-step ahead forecasts for electricity prices using NARX: a new approach, a critical analysis of one-step ahead forecasts**. *Energy Conversion and Management*, v. 50, n. 3, p. 739-747, 2009.

ANDERSOR-SPRECHER, R. **Model comparisons and R^2** . *The American Statistician*, v. 48, n. 2, p. 113-117, 1994.

ASSAREH, E.; BEHRANG, M. A.; GHANBARZDEH, A. **Forecasting energy demand in Iran using genetic algorithm (GA) and particle swarm optimization (PSO) methods**. *Energy Sources, Part B: Economics, Planning, and Policy*, v. 7, n. 4, p. 411-422, 2012.

AREEKUL, P.; SENJYU, T.; URASAKI, N.; YONA, A. **Neural-wavelet approach for short term price forecasting in deregulated power market.** Journal of International Council on Electrical Engineering, v. 1, n. 3, p. 331-338, 2011.

AWAD, M. **Approximating i/o data using wavelet neural networks: control the position of mother wavelet.** The International Arab Journal of Information Technology, v. 9, n. 1, p. 22-29, 2012.

AYALA, H. V. H. **Computational intelligence methods applied to nonlinear black-box system identification.** Tese de doutorado do Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, PR, 2016.

BAKKER, B.; HESKES, T. **Clustering ensembles of neural network models.** Neural Networks, v. 16, n. 2, pp. 261-269, 2003.

BARBOSA GROENNER, B. H. **Computação evolucionária e máquinas de comitê na identificação de sistemas não lineares.** Tese de doutorado do Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2009.

BARROW, D. K.; CRONE, S. F.; KOURENTZES, N. **An evaluation of neural network ensembles and model selection for time series prediction.** The IEEE International Joint Conference on Neural Networks (IJCNN), Barcelona, Espanha, pp. 1-8, 2010.

BATES, J. M.; GRANGER, C. W. J. **The combination of forecasts.** Journal of the Operational Research Society, v. 20, n. 4, pp. 451-468, 1969.

BEN TAIEB, S.; BONTEMPI, G.; ATIYA, A. F.; SORJAMAA, A. **A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition.** Expert Systems with Applications, v. 39, n. 8, pp. 7067-7083, 2012.

BILLINGS, S. A. **Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains.** John Wiley & Sons, Reino Unido, 2013.

BILLINGS, S. A.; VOON, W. S. F. **Structure detection and model validity tests in the identification of non-linear systems.** IEEE Proceedings D - Control Theory and Applications, v. 130, n. 4, pp. 193-199, 1983.

BOWMAN, A. W.; AZZALINI, A. **Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations.** Oxford University Press, 1997.

BREIMAN, L. **Bagging Predictors.** Machine Learning, v.24, pp.123-140, 1996.

BREIMAN, L. **Randomizing outputs to increase prediction accuracy.** Technical Report 518, University of California, Statistics Department, Berkeley: 2000.

BREIMAN, L. **Using iterated bagging to debias regressions.** Machine Learning, v. 45, n. 3, pp. 261-277, 2001.

BROCKWELL, P.; DAVIS R. **Introduction to time series and forecasting.** Springer-Verlag New York, Inc., 2nd ed., 2002.

CARUANA, R; NICULESCU-MIZIL, A.; CREW, G.; KSIKES, A. **Ensemble selection from libraries of models.** Proceedings of the 21st International Conference on Machine Learning, pp. 18-23, Banff, Canada, 2004.

CASTRO, P. D.; COELHO, G. P.; CAETANO, M. F.; ZUBEN, F. J. V. **Designing ensembles of fuzzy classification systems: An immune-inspired approach.** Proceedings of the 4th International Conference on Artificial Immune Systems, pp. 469-482, Banff, Canadá, 2005.

CAVALCANTI, G. D. C.; OLIVEIRA, L. S.; MOURA, T. J. M.; CARVALHO, G. V. **Combining diversity measures for ensemble pruning.** Pattern Recognition Letters, v. 74, pp. 38-45, 2016.

CHAPLAIS, F. **A wavelet matlab tutorial.** Disponível em: <http://cas.ensmp.fr/~chaplais/Wavelet_Matlab/>. Acesso em: 24 nov 2015.

CHAUHAN, N.; RAVI, V.; CHANDRA, D. K. **Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks.** Expert Systems with Applications, v. 36, n. 4, p. 7659-7665, 2009.

CHEN, H.; YAO, X. **Regularized negative correlation learning for neural network ensembles.** IEEE Transactions on Neural Networks, v. 20, n. 12, pp. 1962-1979, 2009.

CHEN, L.; CHIANG, H.; DONG, N.; RONG-PENG, L. **Group-based chaos genetic algorithm and non-linear ensemble of neural networks for short-term load forecasting.** IET Generation, Transmission & Distribution, v. 10, pp. 1440-1447.

CHENG, Y.; CHAN, P.; QIU, Z. **Random forest based ensemble system for short term load forecasting**. International Conference on Machine Learning and Cybernetics, pp. 52-56, 2012.

COELHO, G. P.; VON ZUBEN, F. J. **The influence of the pool of candidates on the performance of selection and combination techniques in ensembles**. International Joint Conference on Neural Networks, pp. 10588-10595, 2006.

CONFLITTI, C.; DE MOL, C.; GIANNONE, D. **Optimal combination of survey forecasts**. International Journal of Forecasting, v. 31, n. 4, pp. 1096-1103, 2015.

DAUBECHIES, I. **Ten lectures on wavelets**. 1ª Ed. – Philadelphia, Estados Unidos da América: Editora SIAM, 1993.

DE FELICE, M.; YAO, X. **Short-term load forecasting with neural network ensembles: a comparative study**. IEEE Computational Intelligence Magazine, v. 6, n. 3, pp. 47-56, 2011.

DOMENICONI, C.; YAN, B. **Nearest neighbor ensemble**. 17th International Conference on Pattern Recognition, Cambridge, Reino Unido, pp. 228-231, 2004.

DREZGA, I.; RAHMAN, S. **Short-term load forecasting with local ANN predictors**. IEEE Transactions on Power Systems, v. 14, n. 3, pp. 844-850, 1999.

DUDEK, G. **Short-term load forecasting using random forests**. Advances in intelligent systems and computing, v. 323, pp. 821-828, 2011.

EL-HAWARY, M. E. **The smart grid – state of the art and future trends**. Electric Power Components and Systems, v. 42, n. 3-4, p. 239-250, 2014.

FAN, S.; CHEN, L.; LEE, W. **Short-term load forecasting using comprehensive combination based on multimeteorological information**. IEEE Transactions on Industry Applications, v. 45, n. 4, p. 1460-1466, 2009.

FONSECA, R.; GÓMEZ, P. **Automatic model selection in ensembles for time series forecasting**. IEEE Latin America Transactions, v. 14, n. 8, pp. 3811-3819, 2016.

FRANK, E.; PFAHRINGER, B. **Improving on bagging with input smearing**. Proceedings of the 10th Pacific-Asia Conference - Advances in Knowledge Discovery and Data Mining, Singapura, pp. 97-106, 2006.

GALAR, M.; FERNÁNDEZ, A.; BARRENECHEA, E.; BUSTINCE, H.; HERRERA, F. **Ordering-based pruning for improving the performance of ensembles of classifiers in the framework of imbalanced datasets.** Information Sciences, v. 354, pp. 178-196, 2016.

GAO, R.; TSOUKALAS, L. H. **Neural-wavelet methodology for load forecasting.** Journal of Intelligent and Robotic Systems: Theory and Applications, v. 31, n. 1-3, pp.149-157, 2001.

GIACINTO, G.; ROLI, F.; FUMERA, G. **Design of effective multiple classifier systems by clustering of classifiers.** Proceedings of the 15th International Conference Pattern Recognition, pp. 160-163, Barcelona, Espanha, 2000.

GRAFF, M.; PEÑA, R.; MEDINA, A.; ESCALANTE, H. J. **Wind speed forecasting using a portfolio of forecasters.** Renewable Energy, v. 68, p. 550-559, 2014.

GRANITTO, P. M.; VERDES, P. F. CECCATO, H. A. **Neural network ensembles: evaluation of aggregation algorithms.** Artificial Intelligence, v. 163, pp. 139-162, 2005.

HABER, R.; UBENHAUEN, H. **Structure identification of nonlinear dynamic systems - a survey on input/output approaches.** Automatica, v. 26, n. 4, pp. 651-677, 1990.

HADAVANDI, E.; SHAHRABI, J.; SHAMSHIRBAND, S. **A novel boosted-neural network ensemble for modeling multi-target regression problems.** Engineering Applications of Artificial Intelligence, v. 45, pp. 204-219, 2015.

HANSEN, L. K.; SALAMON, P. **Neural network ensembles.** IEEE Transactions on Pattern Analysis & Machine Intelligence, n. 10, pp. 993-1001, 1990.

HASSAN, S.; KHOSRAVI, A.; JAAFAR, J. **Examining performance of aggregation algorithms for neural network-based electricity demand forecasting.** International Journal of Electrical Power & Energy Systems, v. 64, pp. 1098-1105, 2015.

HO, T. K. **The random subspace method for constructing decision forests.** IEEE transactions on pattern analysis and Machine Intelligence, v. 20, n. 8, pp. 832-844, 1998.

HONG, T.; PINSON, P.; FAN, S. **Global energy forecasting competition 2012.** International Journal of Forecasting, v. 30, pp.357-363, 2014.

HONG, T.; FAN, S. **Probabilistic electric load forecasting: a tutorial review.** International Journal of Forecasting, v. 32, n. 3, pp.914-938, 2015.

HU, L.; TAYLOR, G. **A novel hybrid technique for short-term electricity price forecasting in UK electricity markets.** Journal of International Council on Electrical Engineering, v. 4, n. 2, p. 114-120, 2014.

ISLAM, M. M.; YAO, X.; MURASE, K. **A constructive algorithm for training cooperative neural network ensembles.** IEEE Transactions on Neural Networks, v. 14, n. 4, pp. 820-834, 2003.

KERMANSHAHI, B.; IWAMIYA, H. **Up to year 2020 load forecasting using neural nets.** International Journal of Electrical Power & Energy Systems, v. 24, n. 9, pp. 789-797, 2002.

KHWAJA, A. S.; NAEEM, M.; ANPALAGAN, A.; VENETSANOPOULOS, A.; VENKATESH, B. **Improved short-term load forecasting using bagged neural networks.** Electric Power Systems Research, v. 125, pp. 109-115, 2015.

KLEIN, C. E. **Identificação não linear multivariável aplicada no ambiente automotivo.** Dissertação de mestrado do Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Pontifícia Universidade Católica do Paraná, Curitiba, PR, 2011.

KOURENTZES, N.; BARROW, D. K.; CRONE, S. F. **Neural network ensemble operators for time series forecasting.** Expert Systems with Applications, v. 41, n. 9, pp. 4235-4244, 2014.

KUNCHEVA, L. I. **Combining pattern classifiers: methods and algorithms.** 2^a Ed. Hoboken, New Jersey: Editora John Wiley & Sons, Inc., 2014.

LAZAREVIC, A.; OBRADOVIC, Z. **Effective pruning of neural network classifiers.** International Joint Conference on Neural Networks, Washington, EUA, pp. 796-801, 2001.

LEE, D.; BALDICK, R. **Short-term wind power ensemble prediction based on Gaussian processes and neural networks.** IEEE Transactions on Smart Grid, v. 5, n. 1, pp. 501-510, 2014.

LEVENBERG, K. **A method for the solution of certain non-linear problems in least squares.** Quarterly of applied mathematics, v. 2, n. 2, pp. 164-168, 1944.

LI, S.; WANG, P.; GOEL, L. **A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection.** IEEE Transactions on Power Systems, v. 31, n. 3, pp. 1788-1798, 2016.

LI, S. Z.; ZHOU, Z. H. **Selective ensemble under regularization framework.** Proceedings of the 8th International Workshop Multiple Classifier Systems, pp. 293-303, Reykjavik, Islândia, 2009.

LIAO, G. **Hybrid improved differential evolution and wavelet neural network with load forecasting problem of air conditioning.** International Journal of Electrical Power & Energy Systems, v. 61, pp. 673-682, 2014.

LIMA, C. A. M. **Comitê de máquinas: uma abordagem unificada empregando máquinas de vetores-suporte.** Tese de doutorado do Programa de Pós-Graduação da faculdade de engenharia elétrica e de computação, Universidade Estadual de Campinas, Campinas, SP, 2004.

LIU, Y.; YAO, X. **Ensemble learning via negative correlation.** Neural Networks, v.12, n.10, pp. 1399-1404, 1999.

LIU, Y.; YAO, X.; HIGUCHI, T. **Evolutionary ensembles with negative correlation learning.** IEEE Transactions on Evolutionary Computation, v. 4, n. 4, pp. 380-387, 2000.

MARGINEANTU, D.; DIETTERICH, T. D. **Pruning adaptive boosting.** Proceedings of the Fourteenth International Conference on Machine Learning, pp. 211-218, San Francisco, 1997.

MARTÍNEZ-MUÑOZ, G.; SUÁREZ, A. **Aggregation ordering in bagging.** Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria, 2004.

MARTÍNEZ-MUÑOZ, G.; SUÁREZ, A. **Pruning in ordered bagging ensembles.** Proceedings of the 23rd International Conference on Machine Learning, pp. 609-616, Pittsburgh, EUA, 2006.

MARTÍNEZ-MUÑOZ, G.; SUÁREZ, A. **Using boosting to prune bagging ensembles.** Pattern Recognition Letters, v. 28, no.1, pp.156-165, 2007.

MCKAY, R.; ABBASS, H. **Anti-correlation: a diversity promoting mechanisms in ensemble learning.** The Australian Journal of Intelligent Information Processing Systems, v. 3, n. 4, pp. 139-149, 2001.

MERZ, C. J.; PAZZANI, M. J. **Combining neural network regression estimates with regularized linear weights.** Proceedings of Advances in Neural Information Processing Systems 9, Denver, EUA, 1996.

MERZ, C. J.; PAZZANI, M. J. **A principal components approach to combining regression estimates.** Machine Learning, v. 36, pp. 9-32, 1999.

MOREIRA, J. M.; SOUSA, J. F.; JORGE, A. M.; SOARES, C. F. **An ensemble regression approach for bus trip time prediction.** Meeting of the Euro Working Group on Transportation, pp. 317-321, 2006.

MOREIRA, J. M.; SOUSA, J. F.; JORGE, A. M.; SOARES, C. F. **Ensemble approaches for regression: a survey.** ACM Computing Surveys, v. 45, n. 1, pp. 1-40, 2012.

NADTOKA, I. I.; BALASIM, M. A. **Mathematical modelling and short-term forecasting of electricity consumption of the power system, with due account of air temperature and natural illumination, based on support vector machine and particle swarm.** Procedia Engineering, v. 129, pp. 657-663, 2015.

NOCEDAL, J.; WRIGHT, S. J. **Numerical optimization.** 1 ed. Nova York: Editora Springer-Verlag New York, Inc., 1999.

NOVOTARSKI, J.; LIU, B.; WERON, R.; HONG, T. **Improving short term load forecast accuracy via combining sister forecasts.** Energy, v. 98, pp. 40-49, 2016.

OPITZ, D. W. **Feature selection for ensembles.** Proceedings of the Sixteenth AAAI Conference on Artificial Intelligence, Orlando, EUA, 1999.

OPITZ, D. W.; SHAVLIK, J. W. **Actively searching for an effective neural network ensemble.** Connection Science, v. 8, pp. 337-354, 1996a.

OPITZ, D. W.; SHAVLIK, J. W. **Generating accurate and diverse members of a neural-network ensemble.** Advances in Neural Information Processing Systems, v. 8, pp. 535-541, 1996b.

ORTIZ-BOYER, D.; HERVÁS-MARTÍNEZ, C.; GARCÍA-PEDRAJAS, N. **CIXL2: Crossover operator for evolutionary algorithms based on population features.** Journal of Artificial Intelligence Research, v. 24, pp. 1-48, 2005.

PAPADOPOULOS, S.; KARAKATSANIS, I. **Short-term electricity load forecasting using time series and ensemble learning methods.** IEEE Power and Energy Conference, Champaign, IL, USA, pp. 1-6, 2015.

PARTRIDGE, D.; YATES, W. B. **Engineering multiversion neural-net systems.** Neural Computation, v. 8. n. 4, pp. 869-893, 1996.

PERRONE, M. P.; COOPER, L. N. **When networks disagree: ensemble methods for hybrid neural network.** Neural Networks for Speech and Image Processing, Ed. Chapman-Hall, pp.126-142, 1993.

PETROVA, S. S.; SOLOV'EV, A. D. **The origin of the method of steepest descent.** Historia mathematica, v. 24, pp. 361-375, 1997.

PUURONEN, S.; TERZIYAN, V.; TSYMBAL, A. **A dynamic integration algorithm for an ensemble of classifiers.** Proceedings of the 11th International Symposium on Foundations of Intelligent Systems, Varsóvia, Polônia, v. 1609, pp. 592-600, 1999.

RANA, M.; KOPRINSKA, I. **Forecasting electricity load with advanced wavelet neural networks.** Neurocomputing, v. 182, pp.118-132, 2016.

RAVIV, Y.; INTRATOR, N. **Bootstrapping with noise: an effective regularization technique.** Connection Science, v. 8, n. 3 e 4, pp. 355-372, 1996.

RAZA, M. Q.; KHOSRAVI, A. **A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings.** Renewable and Sustainable Energy Reviews, v. 50, pp.1352-1372, 2015.

RIBEIRO, G. T.; GRITTI, M.; AYALA, H. V. H.; MARIANI, V. C.; COELHO, L. S. **Short-term load forecasting using wavenet ensemble approaches.** IEEE World Congress on Computational Intelligence (IEEE WCCI), International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, v. 1, pp. 1-8, 2016.

RODRIGUEZ, J. J.; KUNCHEVA, L. I. **Rotation forest: a new classifier ensemble method.** IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 28, n. 10, 2006.

ROKACH, L. **Pattern classification using ensemble methods.** Series in Machine Perception and Artificial Intelligence, v. 75, 2010.

ROONEY, N.; PATTERSON, D.; ANAND, S.; TSYMBAL, A. **Dynamic integration of regression models**. International Workshop on Multiclassifier Systems, vol. LNCS 3181, pp. 164-173, 2004.

ROSEN, B. E. **Ensemble learning using decorrelated neural networks**. Connection Science, v. 8, n. 3, pp. 373-384, 1996.

RUTA, D.; GABRYS, B. **Analysis of the correlation between majority voting error and the diversity measures in multiple classifier systems**. Soft Computing and Intelligent Systems for Industry: Proceedings and Scientific Program: Fourth International ICSC Symposium, Paisley, Escócia, p. 50, 2001.

SCHAPIRE, R. E. **The strength of weak learnability**. Machine Learning, v. 5, n. 2, pp.197-227, 1990.

SHUNMUGAPRIYA, P.; KANMANI, S. **Optimization of stacking ensemble configurations through Artificial Bee Colony algorithm**. Swarm and Evolutionary Computation, v. 12, pp. 24-32, 2013.

SOARES, S. G.; ARAÚJO, R. **An adaptive ensemble of on-line extreme learning machines with variable forgetting factor for dynamic system prediction**. Neurocomputing, v. 171, pp. 693-707, 2016.

SUGANTHI, L.; SAMUEL, A. A. **Energy models for demand forecasting - a review**. Renewable and Sustainable Energy Reviews, v. 16, n. 2, pp. 1223-1240, 2012.

THUILLARD, M. **A review of wavelet networks, wavenets, fuzzy wavenets and their applications**. Advances in Computational Intelligence and Learning, pp. 14-15, 2002.

TODOROVSKI, L.; DZEROSKI, S. **Combining classifiers with meta decision trees**. Machine Learning, v. 50, pp. 223-249, 2003.

TRESP, V.; TANIGUCHI, M. **Combining estimators using non-constant weighting functions**. Advances in Neural Information Processing Systems, pp. 419-426, 1995.

TSANG, I. W.; KOCSOR, A.; KWOK, J. T. **Diversified SVM ensembles for large data sets**. Proceedings of the 17th European Conference on Machine Learning, Berlim, Alemanha, pp. 18-22, 2006.

TUMER, K.; GHOSH, J. **Error correlation and error reduction in ensemble classifiers**. Connection Science, v. 8, n. 3-4, pp. 385-404, 1996.

VILLANUEVA, W. J. P. **Comitê de máquinas em predição de séries**. Dissertação de mestrado do Programa de Pós-Graduação da faculdade de engenharia elétrica e de computação, Universidade Estadual de Campinas, Campinas, SP, 2006.

WANG, H.; FAN, W.; YU, P. S.; HAN, J. **Mining concept-drifting data streams using ensemble classifiers**. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '03, Washington, EUA, v. 2, n. 1, pp. 226-235, 2003.

WICHARD, J. D.; ORGOZALEK, M. **Time series prediction with ensemble models**. IEEE International Conference on Neural Networks, Budapest, v. 2, n. 4, pp. 1625-1630, 2004.

WOLPERT, D. H. **Stacked generalization**. Neural Networks, v. 5, n. 2, pp. 241-259, 1992.

WOLPERT, D. H.; MACREADY, W. G. **No free lunch theorems for optimization**. IEEE Transactions on Evolutionary Computation, v. 1, n. 1, pp. 67-82, 1997.

YANKOV, D.; DECOSTE, D.; KEOGH, E. **Ensembles of nearest neighbor forecasts**. European Conference on Machine Learning, Berlim, Alemanha, v. LNAI 4212, pp. 545-556, 2006.

YKHLEF, H.; BOUCHAFFRA, D. **Induced subgraph game for ensemble selection**. IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), Vietri sul Mare, Itália, pp. 636-643, 2015.

YU, ZHIWEN.; W. DAXING; YOU, J.; WONG, H.; WU, S.; ZHANG, J.; HAN, G. **Progressive subspace ensemble learning**. Pattern Recognition, v. 60, pp. 692-705, 2016.

ZENOBI, G.; CUNNINGHAM, P. **Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error**. Technical Report 2001-11, Trinity College Dublin, Department of Computer Science, Dublin: 2001.

ZHANG, Q.; BENVENISTE, A. **Wavelet networks**. IEEE Transactions on Neural Networks, v. 3, n. 6, pp. 889-898, 1992.

ZHANG, X.; WANG, S.; SHAN, T.; JIAO, L. **Selective SVMs ensemble driven by immune clonal algorithm.** Proceedings of the EvoWorkshop, pp. 325-333, Lausanne, Suíça, 2005.

ZHOU, Z.; WU, J.; TANG, W. **Ensembling neural networks: many could be better than all.** Artificial intelligence, v. 137, n. 1, pp. 239-263, 2002.

ZHOU, Z. **Ensemble methods: foundations and algorithms.** Editora CRC, Boca Raton, Estados Unidos da América, 2012.