

# Comparação de Estratégias de Previsão de Carga Elétrica Utilizando Support Vector Machine

Leandra de Carvalho Nogueira, Ricardo Menezes Salgado

leandra.carvalho@bcc.unifal-mg.edu.br

ricardo@bcc.unifal-mg.edu.br

**Resumo**— A previsão da demanda de carga elétrica é uma tarefa fundamental para um bom funcionamento dos sistemas elétricos. Em um mercado concorrido, para que uma empresa consiga fornecer energia de forma contínua, segura e econômica, esta necessita de técnicas que estimem com máxima precisão o nível de demanda, possibilitando encontrar um ponto ideal de operação que garanta o equilíbrio entre a energia gerada e a consumida. Este trabalho propõe duas estratégias de previsão de carga elétrica, uma baseada na previsão de cada hora individualmente e a outra na linearização dos dados para obter a curva diária através de um único modelo. Os resultados mostraram que as séries com os menores coeficientes de variação apresentaram os menores erros de previsão. A utilização de uma função de autocorrelação não diminuiu os erros de nenhuma das estratégias de previsão propostas e a metodologia de previsão 1 passo a frente se adaptou melhor aos dados trabalhados quando comparada a metodologia de até  $n$  passos a frente.

**Abstract**— The demand forecasting of electric charge is a fundamental task for a proper functioning of electrical systems. In a competitive market, to a company that can supply continuous power, safely and economically, this requires techniques that estimate with maximum precision the level of demand, enabling find an optimal point of operation that ensure the balance between energy generated and consumed. This paper proposes two strategies of electric load forecasting, an approach based on prediction individually each other in time and linearization data for the daily curve through a single model. The results showed, as expected, that the tamer series showed the lowest prediction errors. The use of an autocorrelation function did not decrease any errors in the proposed prediction strategies and methodology for one step ahead prediction is better adapted to the data when compared to the methodology worked up  $n$  steps forward.

**Palavras-chave**— *Previsão de Carga; Support Vector Machines; Modelos de previsão; Inteligência Artificial; Sistema de Potência*

## I. INTRODUÇÃO

A energia elétrica é um bem indispensável à existência da sociedade humana atual. Ela é um insumo básico para residências, comércios, indústrias, bancos, hospitais, empresas, órgãos governamentais e demais entidades em nossa sociedade.

A energia elétrica possui características que a diferencia dos demais insumos industriais. Ela precisa ser gerada simultaneamente ao seu consumo, não pode ser armazenada pelos seus consumidores, tampouco transportada pelos meios usuais de transporte e sua qualidade depende tanto das

empresas, que a produzem, transmitem e distribuem, como dos seus consumidores [1].

O mercado de energia é segmentado em três atividades distintas, mas fortemente acopladas: geração, transmissão e distribuição [2]. Na fase de geração, a energia é produzida efetivamente. Na fase de transmissão, a energia produzida é conduzida, através de linhas de transmissão, para uma fonte de distribuição; e por fim, a fase de distribuição consiste em fazer com que a energia alcance seus consumidores.

Embora a forma como os governos conduzem seus mercados de energia possa variar de um país para outro, observa-se certa tendência em dividir toda a cadeia de fornecimento em pequenas concessionárias de geração e transmissão. Estas concessionárias utilizam leilões de energia para comprar ou vender energia no mercado atacadista [3].

Ao se observar o comportamento do mercado consumidor de energia elétrica, constata-se um crescimento no nível de exigência quanto a qualidade e continuidade do fornecimento de energia. Níveis de oscilação na tensão e de interrupção de fornecimento que antes eram considerados normais, não são atualmente tolerados. A isto se soma a redução da disponibilidade de recursos energéticos renováveis e o acirramento da competição entre as concessionárias de energia elétrica, gerado pela desregulamentação do setor. Estes fatores geram para uma concessionária de eletricidade, a necessidade de reduzir seus custos operacionais e de exercer maior controle sobre seu sistema de potência<sup>1</sup>.

Com o intuito de conseguir êxito neste cenário competitivo, as concessionárias utilizam de diversos recursos técnicos, tal como a predição de carga, para auxiliá-las em suas tomadas de decisões. [4].

Se a predição não for realizada com máxima precisão, as concessionárias podem sofrer severas perdas financeiras. [5], além de sujeitarem os consumidores a problemas de fornecimento [6] [4]. Segundo [7], para uma concessionária de eletricidade do Reino Unido, um incremento de 1% no erro da previsão causa um aumento nos custos operacionais anuais em 10 milhões de libras.

Previsões com alto grau de precisão possibilitam otimizar a produção, reduzir os custos de produção e melhorar a programação para as transferências de energia entre partes do sistema elétrico de uma concessionária ou entre sistemas

---

<sup>1</sup> É o sistema de geração, transmissão e distribuição de energia elétrica, constituído pelas usinas geradoras, linhas de transmissão e subestações.

elétricos de companhias distintas. Um sistema de previsão de cargas preciso é capaz de prover segurança, confiabilidade e economia para a operação de um sistema de potência [8] [9] [10].

A energia elétrica é um bem necessário ao desenvolvimento econômico de um país, então a meta é fazer com que seu uso se dê de forma otimizada e consciente. Do mesmo modo, existe a preocupação em garantir o atendimento das reais necessidades da demanda de carga. Para atender ambas as partes, é necessário que haja um planejamento que procure reduzir ao máximo as perdas e riscos e aumentar a segurança da energia elétrica gerada.

De acordo com [11], o conhecimento futuro do perfil das cargas elétricas é fator essencial no planejamento e operação dos sistemas elétricos de potência.

Segundo [12], o resultado deste planejamento deve prover uma sequência de decisões, de forma a diminuir o custo de operação e garantir o atendimento da demanda, para que as fases de geração, transmissão e distribuição operem em harmonia dentro de seus limites operacionais.

A eficiência deste planejamento reside no fato de encontrar um ponto ideal de operação que irá promover um equilíbrio entre a energia gerada e a consumida. Todo este processo de planejamento envolve uma quantidade considerável de variáveis, exigindo a análise de diversos parâmetros, o que acaba aumentando a sua complexidade.

[13] afirma, que para o caso brasileiro, o setor elétrico utiliza-se das seguintes etapas para implementar o planejamento do parque gerador/transmissor de energia elétrica: planejamento da operação e programação da operação. Na etapa de planejamento da operação, procura-se definir o comportamento da operação do sistema para um horizonte de até cinco passos à frente. Já na etapa de programação da operação, procura-se conciliar a operação do sistema hidráulico e elétrico em horizontes que vão de alguns dias a semanas à frente. Esta operação é muito próxima da operação em tempo real, fazendo com que as restrições elétricas sejam representadas na programação da mesma.

Após o sistema ser projetado para suportar o transporte de uma determinada quantidade de carga, fica evidente que o conhecimento da carga que o sistema terá que transmitir é essencial para que se possa efetuar um planejamento operacional sem causar danos à rede.

Percebe-se que o desequilíbrio, tanto da energia gerada em excesso quanto da energia gerada em quantidade menor que o necessário, pode causar problemas de sobre tensão e sub tensão no sistema. Existem duas soluções possíveis para tentar amenizar este problema ou tentar promover um equilíbrio entre escassez e excesso de energia: a reprogramação das geradoras ou a transferência de energia através dos intercâmbios entre os sistemas. Entretanto, estas atividades demandam certo intervalo de tempo para serem efetivadas e daí surge e motivação de se realizar uma boa previsão [14].

Sabendo que o ideal seria que toda a energia produzida fosse consumida na sua totalidade, sem desperdícios ou escassez, é de fundamental importância que se descubra o comportamento da carga futura com máxima precisão para

que a operação de todo o sistema elétrico atenda apenas a quantidade de demanda necessária.

O estudo e análise de carga elétrica acabam se tornando uma exigência fundamental durante a operação do sistema como um todo. Implementar novas técnicas que procurem garantir um fornecimento de energia com qualidade é um dos requisitos que as empresas do setor elétrico tem que realizar.

Em concordância com este objetivo, o conhecimento do comportamento da carga futura é o primeiro pré-requisito para um planejamento seguro e econômico do sistema de energia elétrica. Aumentar a segurança e a economia da operação de sistemas são as motivações para a realização de previsões de carga mais confiáveis nos sistemas elétricos de potência [15].

Visando melhorias neste contexto, novas ferramentas de otimização e controle são estudadas, desenvolvidas e aplicadas, fornecendo resultados que proporcionam redução de determinados custos relacionados. Uma das técnicas mais utilizadas é a análise e previsão de séries temporais. Esta técnica visa a obter valores futuros de séries de vazões, consumo, demanda, entre outros, através do histórico prévio destas séries.

Existem diversos modelos que são utilizados para previsão de séries temporais, destacando-se os modelos estatísticos e os modelos baseados em inteligência computacional. Os modelos estatísticos são utilizados para compreender o processo de formação dos dados temporais, e assim descrever efetivamente o comportamento da série. Os modelos baseados em inteligência computacional não dependem de complexos modelos matemáticos e tem como vantagem obter uma melhor resposta a não linearidade das séries temporais.

Na literatura existem diversas técnicas sendo estudadas e desenvolvidas visando descobrir um modelo que atenda a grande maioria dos casos para a solução do problema de previsão de carga elétrica.

O trabalho de [16] utilizou a biblioteca LibSVM para as previsões anuais de carga no horizonte de dez anos e para previsões mensais de carga. Os autores afirmam que a LibSVM apresentou excelente desempenho durante as previsões. Eles também perceberam que havia necessidade dos dados da série trabalhada serem maiores e a correlação entre temperatura e carga ser não significante para se obter melhores previsões.

Considerando o trabalho de [17], a autora desenvolveu uma metodologia de previsão de carga multinodal à curto prazo (24 horas à frente), que pode ser entendida como um sistema inteligente que considera vários pontos de rede elétrica durante a realização da previsão. No trabalho foi utilizada RNA composta por vários módulos variando a função de ativação. O erro percentual absoluto médio dos resultados ficou em torno de 2% e a função de ativação que apresentou o melhor resultado foi a secante hiperbólica.

Já no trabalho de [18], os autores utilizam como componentes individuais de previsão as Redes Neurais, as *Support Vector Machine*, as redes de Elman e as redes de Kohonen, objetivando estimar a carga futura 24 horas à frente. Eles ainda utilizam uma forma de combinar os resultados denominada BSS (*Blind Source Separation*), em que é feita

uma análise das séries temporais previstas pelas componentes, separando as que apresentam ruídos e somente as que apresentam os resultados mais significantes é que participam da reconstrução do resultado final.

Outro trabalho que realiza previsão de carga elétrica é o de [19]. Os autores realizam a previsão horária de carga utilizando duas redes neurais idênticas. Os resultados obtidos pelas redes são combinados através da média entre eles e são utilizados para realimentar as previsões subseqüentes.

Nos trabalhos de [20] e [21], os autores utilizam sistemas híbridos para realizar a previsão de cargas a curto prazo. O primeiro faz uso da PSO (*particle swarm optimization*) para realizar o treinamento de uma Rede Neural Artificial e o segundo utiliza redes de funções de base radial (RBF) e o ANFIS, que é um sistema que combina lógica *fuzzy* com Redes Neurais Artificiais, para realizar o treinamento.

O trabalho de [22], utiliza modelos baseados em kernel para previsão de carga a curto prazo. Os autores investigam a minimização de limites superiores do erro de generalização de *Support Vector Machines* (SVMs) e máquinas de vetores relevantes. Foram utilizados três bases de dados que apresentam dados horários de carga e de temperatura. Com exceção da primeira base, a minimização do limite superior do erro de generalização produziu melhorias de desempenho para as SVMs em termos de capacidade preditiva.

Analisando o trabalho de [23] percebe-se que o autor aplicou métodos de previsão a longo prazo para previsão de consumo de energia elétrica em séries com tendência e ciclos. O autor utiliza os modelos ARIMA (*AutoRegressive Integrated Moving Average*), NARIMA (*Nonlinear AutoRegressive Integrated Moving Average*), RNF (Rede *Neuro-Fuzzy*) e a RNA (Rede Neural Artificial) e realiza previsões de 60 passos a frente. Para a realização da pesquisa foram utilizadas duas séries temporais: uma de New England e uma da Cemig. Os melhores resultados para ambas as séries foram os obtidos pela RNA, sendo 2,383% e 1,503%, respectivamente.

O trabalho de [24] foi desenvolver um modelo de previsão de carga diária ativa, em base horária, por barramento. O autor abordou dois tipos de metodologias: uma de previsão individual (MPI), onde cada barramento é tratado de forma isolada e outra na qual a previsão é feita uma única vez para um dado conjunto de barramentos (MPD). Para realizar a previsão foi utilizada uma rede neural de múltiplas camadas do tipo *perceptron* com 1 camada intermediária. Os resultados mostraram que o MPI apresentou melhores ajustes na maioria dos barramentos, no entanto a metodologia de previsão agregada foi eficiente comparada ao MPI, apresentando resultados aceitáveis e um número menor de previsões.

Já o trabalho de [25], propõe uma metodologia de previsão de carga elétrica à curto prazo que envolve a combinação de vários previsores distintos. A combinação envolve as componentes de programação genética, RNAs e regressão linear múltipla. Para execução dos experimentos são considerados dados de medição obtidos do sistema elétrico brasileiro. Dos sete dias testados somente em um, o *ensemble* não superou a melhor componente individual.

O trabalho de [26] propõe uma nova técnica para previsão de carga a curto prazo baseada na transformada *wavelet* discreta. O objetivo dos autores foi desenvolver um previsor neural de carga mais robusto. Para testar a metodologia proposta, eles utilizaram dados de carga de uma concessionária de energia elétrica norte-americana. Os autores testaram 4 modelos de previsão, sendo que o que melhor se adaptou foi o que combina normalização, padronização e análise multirresolução via *wavelets*, com erros em aproximadamente 3%.

Neste contexto, a metodologia utilizada neste trabalho é baseada na comparação de diferentes modelos de previsão a fim de verificar qual deles melhor se adapta a série de energia trabalhada. Os modelos e as formas de previsão são variados com o intuito de verificar qual a influência desta mudança no erro de previsão.

Este artigo está organizado da seguinte forma: na seção II é apresentada a fundamentação teórica das técnicas utilizadas neste trabalho, na seção III é exposta a metodologia utilizada para o desenvolvimento da pesquisa, na seção IV são apresentadas análises dos dados trabalhados, na seção V são mostrados os resultados obtidos pelas estratégias de previsão propostas e finalmente na seção VI são feitas as considerações finais.

## II. FUNDAMENTAÇÃO TEÓRICA

Antes de apresentar a metodologia abordada, serão apresentados os aspectos teóricos utilizados para a execução e desenvolvimento da pesquisa.

### A. Séries temporais

Uma série temporal, também denominada série histórica, é uma sequência de dados obtidos em intervalos regulares de tempo durante um período específico. [27] [28]. Este conjunto pode ser obtido através de observações periódicas do evento de interesse.

Se a série histórica for denominada como  $Z$ , o valor da série no momento  $t$  pode ser escrito como  $Z_t$  ( $t = 1, 2, \dots, n$ ). Denomina-se trajetória de um processo, a curva obtida no gráfico da série histórica e o conjunto de todas as possíveis trajetórias é denominado como um processo estocástico. Considera-se que uma série temporal é uma amostra deste processo.

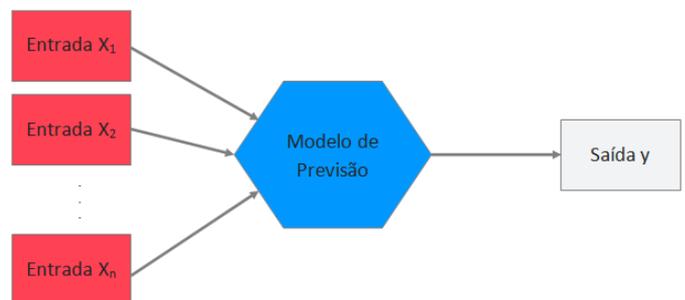


Figura 1 - Modelo clássico de Previsão

A Fig. 1 retrata o esquema clássico de um modelo de previsão de séries temporais, com várias entradas em instantes anteriores e uma saída que representa a previsão no instante  $t$ .

Por possuir um comportamento não linear, o comportamento da série fica difícil de ser avaliado e previsto. Além do fato que a complexidade do problema pode variar de acordo com o período e o número de séries a serem previstas.

Analisando o problema de previsão de carga elétrica, pode-se classificá-lo como um problema de previsão de séries temporais. Neste processo, o histórico de dados de consumo de carga é utilizado para compor e representar as séries em questão.

### B. Support Vector Machines (SVM)

O aprendizado de máquina (AM) é um conjunto de técnicas computacionais que tem por objetivo a criação de sistemas capazes de adquirir e organizar o conhecimento de forma automática. [29].

Um das técnicas do AM é o aprendizado supervisionado que consiste em criar uma função através de um conjunto de treinamento. [29]. Esse conjunto possui pares de objetos de entrada (tipicamente vetores) e saídas desejadas, que podem ser um número real (para os casos de regressão) ou um rótulo de uma classe (para os casos de classificação).

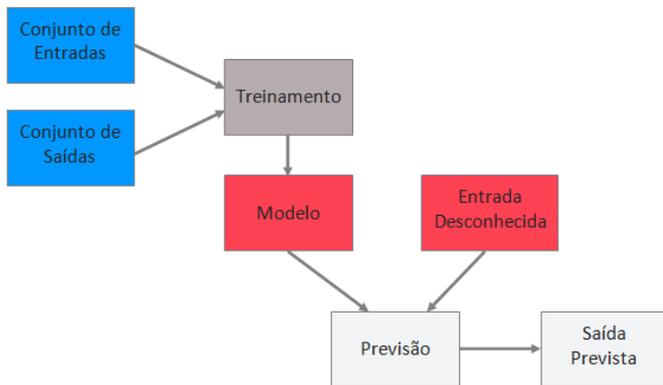


Figura 2 - Fluxograma de aprendizado e previsão dos algoritmos do tipo Aprendizado Supervisionado

O objetivo do aprendizado supervisionado (AS) é utilizar a função criada para prever o valor de saída (resultado) através dos dados de entrada. A Fig. 2 mostra os passos de aprendizado e previsão dos algoritmos baseados em aprendizado supervisionado AS.

Os algoritmos principais dessa técnica são: RNAs, algoritmo *Nearest Neighbor* (NN), árvores de decisão e a máquina de vetores suporte (SVM), entre outros.

O SVM pode ser entendido como um conjunto de técnicas utilizadas tanto para classificação (*Support Vector Classification – SVC*) quanto para regressão (*Support Vector Regression – SVR*). Utiliza-se da minimização estrutural de risco (*Structural Risk Minimization – ERM*) que é utilizado pelas redes neurais mais convencionais. [30]. O uso do conceito de ERM possibilita que a SVM tenha uma melhor generalização e por consequência tenha um bom desempenho na aprendizagem estatística.

Para a construção do conhecimento, a técnica de aprendizado SVM utiliza como conceito chave a minimização do risco estrutural, que é baseada na dimensão de Vapnik-

Chevonenkis. É importante visar que a técnica SVM é dividida em duas vertentes, pois embora tenha sido criada para atender casos em que as saídas são números discretos (SVC), ela foi modificada para também atender casos em que as saídas são números contínuos (SVR).

Entre suas características principais destacam-se a alta capacidade de generalização e a robustez da máquina de vetores suporte quando confrontada com dados de grande dimensão, uma vez que não tende a ficar super (*overfitting*) ou sub (*underfitting*) ajustada, ao contrário de outras técnicas de aprendizado. [31].

Embora possua muitas características positivas, o SVM possui algumas limitações que prejudicam a sua usabilidade: velocidade de treinamento menor em relação a outras técnicas, alta complexidade computacional, sensibilidade e dificuldade na escolha dos parâmetros e dificuldade de interpretação do conhecimento adquirido.

Entretanto, diversos estudos foram feitos para diminuir essas limitações, o que somado com a alta capacidade de generalização e a sua robustez faz do SVM uma das técnicas de aprendizado mais exploradas e estudadas atualmente. [32].

### C. Dimensão de Vapnik-Chervonenkis (VC)

A dimensão Vapnik-Chervonenkis ou dimensão VC é medida da capacidade de classificação de um grupo de funções indicadoras calculadas previamente por uma máquina de aprendizagem. O seu valor é equivalente ao número máximo de exemplos de treinamento que podem ser aprendidos sem erros.

O valor da dimensão VC é  $n + 1$ , sendo  $n$  a dimensão do espaço vetorial em questão. Por exemplo,  $VC = 2$  quando o problema pode ser separado por uma reta,  $VC = 3$  quando o problema pode ser separado por um plano e assim por diante. A Fig. 3 mostra um exemplo do valor da dimensão. Neste exemplo, o valor VC é 2, uma vez que o problema pode ser dividido por uma reta.

A dimensão VC é utilizada como parâmetro de controle para ajustar um princípio indutivo conhecido como minimização do risco estrutural. [33] [34]. Porém calcular a dimensão VC não é uma tarefa trivial, podendo o valor ser desconhecido ou infinito. [31].

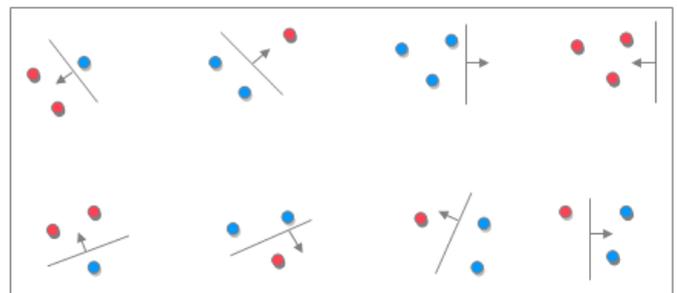


Figura 3 - Dimensão VC=2. Fonte: [35]

### D. Minimização do Risco Estrutural (SRM)

Para obter um melhor desempenho nas previsões é esperado que a função de classificação  $f$  consiga chegar ao menor erro possível durante o treinamento. É importante notar

que o erro nesse caso é o número de predições incorretas de  $f$ , dessa forma o risco empírico é a medida de perda entre a resposta desejada e a resposta obtida (real). A definição matemática do risco empírico é mostrada na Eq. 1, onde  $c(f(x_i), y_i)$  é a função de custo da previsão  $f(x_i)$  tendo  $y_i$  como saída desejada e  $n$  o número de elementos do conjunto de treinamento [31].

$$R_{empirico}(f) = \frac{1}{n} \sum_{i=1}^n c(f(x_i), y_i) \quad (1)$$

Através da dimensão  $VC$  foi desenvolvida uma expressão que indica o limite superior do risco funcional tendo probabilidade  $1 - e$  de ocorrer, sendo  $e$  um valor entre 0 e 1. [36]. Essa expressão pode ser visualizada na Eq. 2.

$$R_{funcional}(f) \leq R_{empirico} + R_{estrutural}(h, N, e) \quad (2)$$

A formulação do risco estrutural pode ser visualizada na Eq. 3, sendo  $h$  a dimensão  $VC$  da classe de funções  $F$  na qual  $f$  pertence,  $N$  o número de amostras de treinamento e  $e$  o parâmetro escolhido.

$$R_{estrutural}(h, N, e) = \sqrt{\frac{h \left( \ln \left( \frac{2N}{h} + 1 \right) - \ln \left( \frac{e}{4} \right) \right)}{N}} \quad (3)$$

O ideal é encontrar um valor que equilibre o risco empírico com o risco estrutural, uma vez que a diminuição do primeiro implica no crescimento do segundo, ou seja, deve-se obter o menor limite possível para o risco funcional.

Um grande problema da minimização do risco estrutural está ligado à complexidade para encontrar o valor exato da dimensão  $VC$  ( $h$  da fórmula), então o valor costuma ser estimado por meio de limites superiores. [37]. A fórmula que mostra esta estimativa pode ser vista na Eq. 4.

$$h \leq 1 + \min \left( n, \left( \frac{r^2}{T^2} \right) \right) \quad (4)$$

Sendo  $T$  a margem de separação entre as classes e  $r$  o menor raio possível de uma hiper-esfera circunscrita aos vetores de treinamento. Outra maneira de minimizar o risco estrutural é a maximização da margem de separação  $T$ , uma vez que  $T$  é inversamente proporcional à dimensão  $VC$ .

#### E. Support Vector Classification (SVC)

Ao realizar a classificação o SVC utiliza o conceito de margem. Para entendê-lo deve-se considerar um hiperplano  $f(x) = (w \cdot x) + b$ . A margem é a menor distância entre os exemplos do conjunto de treinamento e o hiperplano utilizado para separação dessas classes [31]. Ou seja, a margem determina quão bem duas classes podem ser separadas [38].

De maneira mais formal, o SVC procura maximizar a margem utilizando um hiperplano ótimo, ou seja, o hiperplano de separação que possui a maior margem considerando um conjunto de treinamento linearmente separável, o que pode ser obtido através da Eq. 5, considerando como restrição  $y_i(w \cdot x_i + b) - 1 \geq 0, \forall i = 1, \dots, n$ . [31]. Ao minimizar esta equação implementa-se o princípio de minimização do risco estrutural (SRM), que supõe a desigualdade  $\|w\| < A$ , sendo  $A$  um escalar dos números reais [39].

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (5)$$

Outro conceito importante para o SVC é o de vetores suporte. Esses vetores são padrões críticos que sozinho determinam o hiperplano ótimo, enquanto os outros padrões (não críticos) são considerados irrelevantes. [38]. Na Fig. 4 está um exemplo de margem e vetores suporte (são destacados por círculos externos aos padrões).

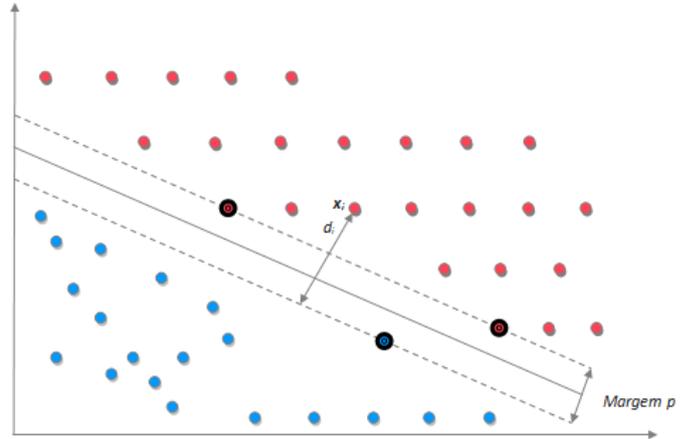


Figura 4 - Margem  $p$  e os vetores de suporte. Fonte: [40]

A predição de uma amostra para classificação binária é realizada pela fórmula da Eq.6, em que  $n$  é a quantidade de elementos do conjunto de treinamento e  $\alpha$  e  $b$  são parâmetros criados pelo SVM para a construção de uma solução otimizada. [41]. Para a classificação multiclasse pode-se recorrer a duas soluções: decomposição do problema para vários subproblemas binários ou a modificação do algoritmo do SVM. Como o segundo método geralmente é custoso computacionalmente, o método mais comumente usado é o de decomposição [31].

$$f(x) = \text{sign}(\sum_{i=1}^n y_i \alpha_i (x \cdot x_i) + b) \quad (6)$$

Quando o problema não é linearmente separável é introduzido o conceito de erro para que o método de separação de hiperplanos possa ser generalizado. Isso é feito através das variáveis de folga  $\xi_i$ , para todo  $i = 1, \dots, n$ , que tornam a solução mais flexível considerando o hiperplano como  $y^i[(w \cdot x^i) + b] \geq 1 - \xi_i$ , sendo  $\xi_i \geq 0$ . Portanto o hiperplano que maximiza a margem separadora pode ser obtido pela minimização da Eq. 7, chamada de margem suave. Segundo [42], o parâmetro  $C$  de regulação deve ser escolhido na maioria dos casos empiricamente e deve refletir a existência de ruído nos dados.

$$\phi(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \quad (7)$$

Esse procedimento é dado pelo teorema de Cover, que diz que dado um conjunto de dados não linear no espaço  $X$ , esse espaço pode ser transformado no espaço de características  $\mathfrak{F}$  que possui uma alta probabilidade de ter os dados linearmente separáveis [31].

Como  $\mathfrak{F}$  pode ter dimensão muito grande e até mesmo infinita, a computação de  $\phi$  pode se tornar inviável. Portanto é necessário modificar a fórmula de previsão, através das

chamadas funções de *Kernel*. Essas funções substituem o  $(x \cdot x_i)$  da Eq. 7, para permitir o crescimento da solução SVC gerando uma melhor classificação.

F. Funções Kernel

Uma função *Kernel* é uma função que recebe como parâmetro  $x_i$  e  $x_j$  do espaço de entradas e calcula o produto escalar desses parâmetros no espaço de características. (Eq.8). Sendo que a função  $\phi(\cdot)$  deve pertencer a um domínio onde seja possível o cálculo do produto interno.

$$K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle \quad (8)$$

Os tipos mais comuns de *Kernels* podem ser vistos na Tab. 1, sendo que em alguns deles é necessária a configuração de parâmetros adicionais.

Tabela 1 - Tipos mais comuns de Kernels. Fonte: [31]

Tipo de Kernel	Informações dos Kernels	
	Função $K(x_i, x_j)$	Índice B
Polinomial	$(\delta(x_i \cdot x_j) + k)^d$	$\delta, K$ e $d$
Gaussiano	$\exp(-\sigma \ x_i - x_j\ ^2)$	$\sigma$
Sigmoidal	$\tanh(\delta(x_i \cdot x_j) + k)$	$\delta$ e $k$

G. Support Vector Regression (SVR)

A diferença da regressão para a classificação está no fato da primeira apresentar muitos e possivelmente parâmetros de saída tendendo ao infinito. Ou seja  $y_i \in R$  ao invés de ser parte de um número finito de categorias. Isso torna a regressão uma generalização da classificação, em que a resposta varia continuamente dentro de uma faixa esperada.

A regressão em SVM é semelhante à classificação, mas utiliza-se uma função de perda. Essa função pode ser alterada dependendo o problema, mas a mais comum é a  $\epsilon$ -insensitiva criada por Vapnik.

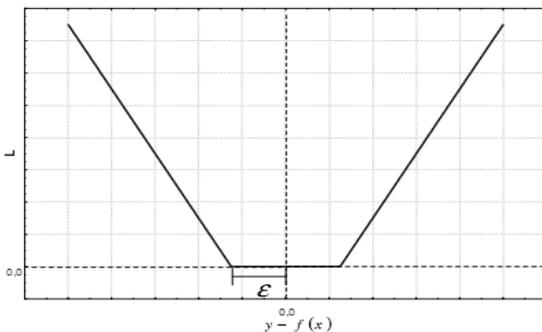


Figura 5 – Função de perda  $\epsilon$ -insensitiva. Fonte:[43]

A Fig. 5 mostra como o erro  $\epsilon$  funciona. Quanto maior o erro, maior é a penalização  $L$ . Se não houver erro, não há penalização. Isso forma uma região de tolerância determinada por  $(-\epsilon, +\epsilon)$  que é chamada de região insensitiva. Entretanto, quando o erro não está dentro desta região é aplicada a penalização em proporção linear ao seu módulo. [43]. Ou seja, o erro não é levado em consideração se ele for menor que  $\epsilon$ ,

porém não será aceito nenhum erro maior que este parâmetro. [44].

O conceito de margem, assim como na classificação, também é empregado. A margem do SVR é uma região em torno de uma curva ajustada a partir dos erros cujos módulos extrapolam  $\epsilon$ . Essa região mostra a região de confiança, como mostrado na Fig. 6.

As previsões são feitas de forma semelhante ao SVC levando em consideração o erro para se obter números contínuos como resultados e do mesmo como no problema da classificação, são utilizadas as funções *Kernel* quando o problema não pode ser separado de forma linear.

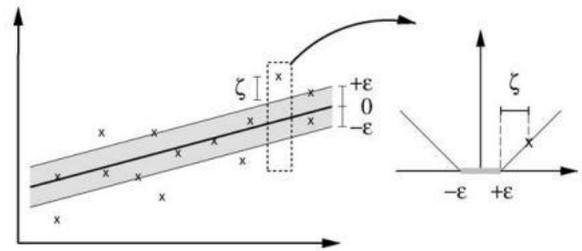


Figura 6 – Conceito de margem no SVR. Fonte:[44]

H. Função de Autocorrelação

Em pesquisas, frequentemente procura-se verificar se existe relação entre duas ou mais variáveis, isto é, saber se as alterações sofridas por uma das variáveis são acompanhadas por alterações nas outras.

Em previsão de séries temporais com utilização de abordagens estatísticas, a função de autocorrelação (AFC) é extremamente útil na obtenção de uma descrição parcial da série [45].

A ACF mede o grau de correlação de uma variável, em um dado instante, consigo mesma, em um instante de tempo anterior. Pode ser definida como a razão entre a autocovariância e a variância para um conjunto de dados. A Eq. 9 determina a correlação entre observações separadas por um atraso  $k$  [46].

$$r_k = \frac{\sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (9)$$

O valor  $r_k$  é denominado coeficiente de autocorrelação em um atraso  $k$  e descreve a autocorrelação entre  $x_i$  e  $x_{i+k}$ . A variável  $N$  representa a quantidade de valores da amostra de dados utilizada,  $\bar{x}$  representa a média dos valores da amostra e  $x_i$  representa a variável que se deseja medir o grau de autocorrelação [46].

Quanto maior o valor de  $r_k$ , mais o grau de autocorrelação entre  $x_i$  e  $x_{i+k}$  [45].

III. METODOLOGIA UTILIZADA

Este trabalho foi desenvolvido utilizando a biblioteca LibSVM [47] que implementa computacionalmente os métodos das *Support Vector Machines* após a configuração dos parâmetros, que serão explicados na Tab 2. Esta biblioteca é consolidada na literatura por estabelecer uma rápida

convergência, alta portabilidade e atualizações frequentes [48].

A Fig. 7 mostra a metodologia utilizada para encontrar a melhor componente para cada uma das séries trabalhadas. As fases são executadas de forma sequencial: na primeira é ajustado e configurado o modelo de previsão, na segunda o modelo é utilizado para prever uma amostra desconhecida e na última fase, a melhor componente é selecionada.

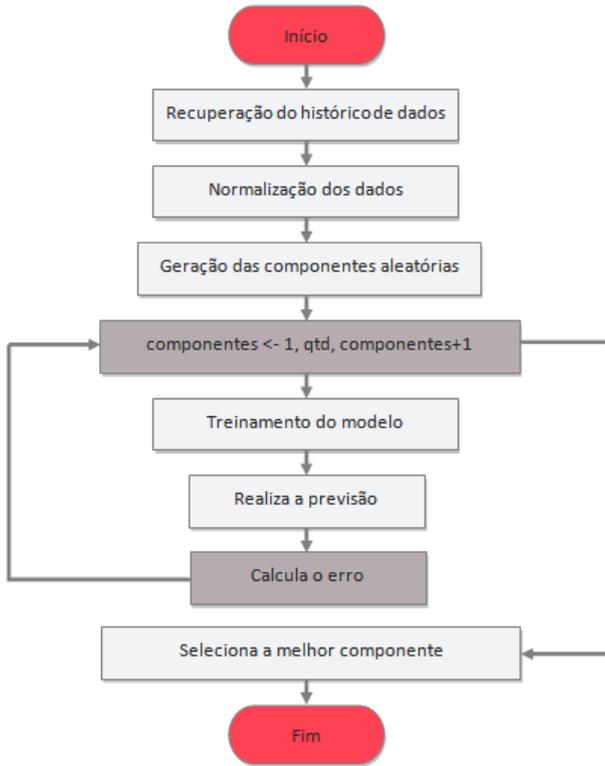


Figura 7 - Fluxograma da metodologia utilizada

O conjunto de dados recuperado foi dividido em dois: a primeira parte foi utilizada para treino e a segunda para teste.

Para a correta utilização da biblioteca e a otimização dos resultados, os dados de carga elétrica foram normalizados através da Eq. 10, onde  $x$  representa o dado a ser transformado,  $\max(x)$  se refere ao maior valor de toda a base de dados e  $x'$  representa o dado transformado.

$$x' = \frac{x}{\max(x)} \quad (10)$$

Para configuração da SVM foram geradas 1000 componentes de previsão com valores aleatórios para sete parâmetros de configuração. Estes parâmetros foram escolhidos de forma aleatória devido ao grande número de possibilidades e intervalos de configurações. Esta combinação entre diferentes parâmetros tem por objetivo encontrar o mais bem adaptado para cada uma das séries de carga trabalhadas.

A Tab. 2 mostra os parâmetros de configuração utilizados para a geração das componentes, o intervalo de variação de cada um e o seu significado.

Tabela 2 - Configuração do parâmetro da SVM

Parâmetro	Informações dos parâmetros de configuração	
	Intervalo	Significado
s	3 ou 4	O tipo de svm. Os dois tipos utilizados para regressão são o 3 (epsilon-SVR) e o 4 (nu-SVR).
d	1,2 ou 3	O valor do grau da função do Kernel escolhido.
g	1 a 100	O valor gamma da função do Kernel escolhido.
c	0.01 a 500	O valor do parâmetro C (custo) do C-SVC, epsilon-SVR e nu-SVR.
p	0.1	O valor de epsilon na função de perda do epsilon-SVR
n	1 a 100	O valor do parâmetro nu de nu-SVC, one-class SVM e nu-SVR.
t	0,1,2 ou 3	O tipo de kernel, os quais podem ser linear (0), polinomial (1), gaussiano (2) ou sigmoidal (3).

Como dito anteriormente, este trabalho tem por finalidade comparar diferentes modelos de previsão a fim de verificar qual deles melhor se adapta a série de energia trabalhada e qual a influencia de cada uma destas estratégias no erro de previsão. Para tal, foram codificados duas estratégias de previsão: o primeiro realiza previsão de cada hora individualmente para a obtenção da curva diária, enquanto o segundo lineariza os dados para obter a curva diária através de um único modelo.

As estratégias, os métodos de previsão (1 passo ou n passos) e a forma de escolher as entradas de cada saída desejada (atrasos ou autocorrelação) foram variados a fim de verificar qual a influência desta mudança no erro de previsão.

Ambos os modelos utilizam os métodos de previsão 1 passo a frente, onde a previsão leva em consideração o histórico recente com cargas conhecidas até o dia anterior a ser previsto e até n passos a frente, onde a previsão é realizada utilizando um histórico de cargas conhecido até uma determinada hora fixa e também toma como base os valores de carga já estimados pelo modelo (previsão sobre previsão).

Para melhor compreensão de como as entradas são formadas, será apresentado um exemplo de construção do par entrada e saída desejada para uma determinada hora de um dia.



Figura 8 - Exemplo de formação de entradas e saídas para previsão de uma determinada hora de uma quinta-feira.

A Fig. 8 apresenta uma previsão para uma determinada hora de uma quinta-feira (destacada em azul). Os valores de carga das quintas-feiras que antecedem a quinta-feira almejada compõem as saídas desejadas e são escolhidas 3 entradas para cada uma destas quintas-feiras (destacadas em rosa).

As entradas de cada saída desejada para cada um dos métodos de previsão (1 passo a frente e até n passos a frente) são escolhidas mediante um atraso de 3,5,7 dias ou um número de dias escolhido por uma função de autocorrelação. Esta função determina os índices dos valores do vetor de entradas que serão utilizados. Para este exemplo foi utilizado um atraso de 3 dias.

Para a utilização da função de autocorrelação foi utilizada uma amostra de 90 dados anteriores para a primeira estratégia e 3650 para a segunda estratégia. Esta quantidade de dados influencia no número de padrões de treinamento e consequentemente na velocidade de previsão do modelo. Os valores utilizados foram escolhidos de forma a deixar o modelo rápido e com boas previsões.

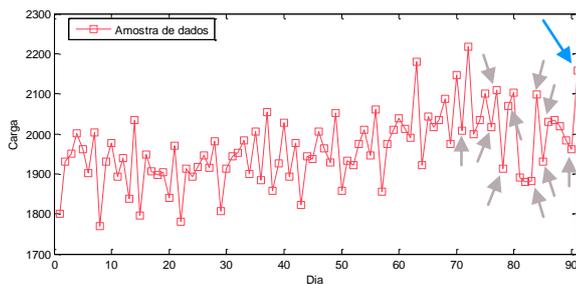


Figura 9 - Exemplo da escolha de índices por uma função de autocorrelação

A Fig. 9 apresenta um exemplo de como a função de autocorrelação escolhe os índices que irão compor as entradas e saídas desejadas. Neste exemplo, os valores de carga autocorrelacionados com a saída almejada (representada pela seta em azul) são representados em cinza. Para esta amostra de 90 valores, 10 valores autocorrelacionados foram encontrados, que são representados, da direita para a esquerda, pelos índices: 89, 85, 84, 83, 82, 78, 77, 76, 75 e 70.

Realizadas as previsões, o critério adotado para escolher a melhor componente é baseado no erro percentual absoluto médio (MAPE), definido na Eq. 11, onde  $X_{previsto}$  representa o valor previsto pela componente e  $X_{real}$  representa o valor real.

$$MAPE = 100 \frac{1}{n} \sum_{i=1}^n \left| \frac{X_{previsto} - X_{real}}{X_{previsto}} \right| \quad (11)$$

Estabelecida a forma de cálculo, a seleção da melhor componente para cada série é em função daquelas que obtiveram o menor MAPE.

#### IV. ANÁLISE DOS DADOS

Visando compreender melhor o comportamento dos dados, foram feitas análises descritivas. A Tab. 3 mostra a média, o desvio padrão, o valor máximo, o valor mínimo e o coeficiente de variação para as séries completas e para cada período que se deseja realizar a previsão.

Tabela 3 - Medidas de posição e variabilidade das séries completas e seus respectivos períodos de previsão.

Período	Medidas de Posição e Variabilidade				
	Média	Desvio Padrão	Valor Máximo	Valor Mínimo	Coefficiente de Variação
Série 1 Completa	1712,8	27,6	2659,6	889,6	0,016
Série 2 Completa	549,4	11,7	781,8	332,1	0,021
Série 3 Completa	1767,0	50,4	2466,1	813,5	0,029
Previsão Série 1	2033,0	24,2	2659,6	1361,4	0,012
Previsão Série 2	590,1	12,5	781,8	377,5	0,021
Previsão Série 3	1874,8	53,6	2466,1	1001,7	0,029

Analisando a Tab. 3 percebe-se que a Série 1 é a que apresenta a menor variabilidade dos dados, tanto na série completa quando no período utilizado para previsão. Nota-se ainda que se tratando da série de previsão, este valor diminuiu em 4%. A Série 2 manteve a mesma variabilidade dos dados completos e dos dados de previsão, o que também ocorre para a Série 3.

Estas informações são importantes, pois espera-se que dados com os menores coeficientes de variação se consigam melhores previsões.

Os históricos utilizados para treinamento e teste dos modelos são formados por dados reais da demanda de carga elétrica de diferentes empresas fornecedoras de energia elétrica. O período do histórico de medição para a Série 1 compreende o intervalo entre 01/01/2006 e 31/10/2011, a Série 2 compreende o intervalo entre 01/01/2007 e 31/03/2012 e a Série 3 entre 01/01/2007 e 31/12/2011. Os dados são apresentados hora a hora.

A Fig. 10 apresenta o período de previsão para a Série 1, considerando a primeira estratégia. Nesta estratégia, as séries apresentam a carga hora a hora para todos os dias pelo período de 1 ano. O período do histórico de previsão para esta série compreende o intervalo entre 31/10/2010 e 31/10/2011.

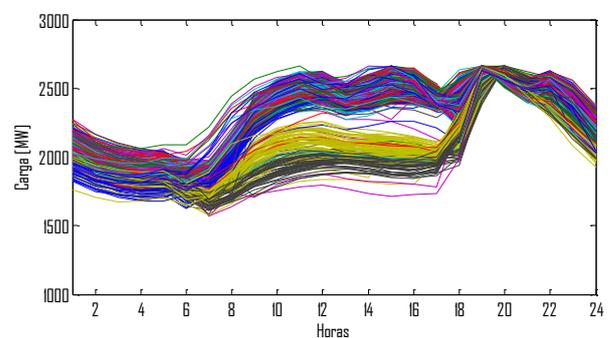


Figura 10 – Período de previsão para a Série 1 utilizando a 1ª estratégia.

A Fig. 11 apresenta este mesmo período, porém considerando a segunda estratégia. Nesta estratégia dados apresentados foram linearizados formando uma única curva diária.

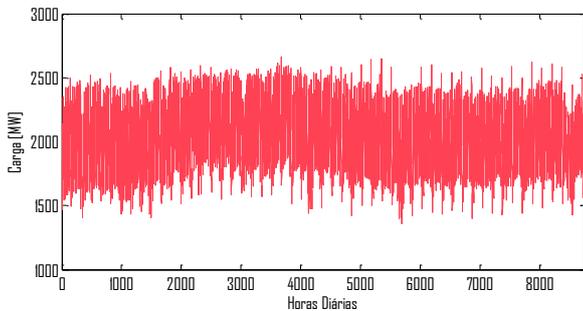


Figura 11 - Período de previsão para a Série 1 utilizando a 2ª estratégia.

A Fig. 12 apresenta o período de previsão para a Série 2, considerando a primeira estratégia. O período do histórico de previsão para esta série compreende o intervalo entre 31/03/2011 e 31/03/2012.

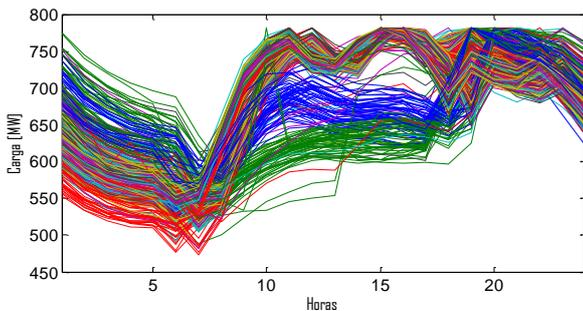


Figura 12 - Período de previsão para a Série 2 utilizando a 1ª estratégia.

A Fig. 13 apresenta este mesmo período, porém considerando a segunda estratégia.

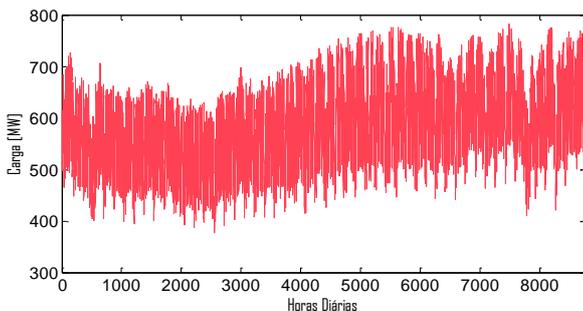


Figura 13 - Período de previsão para a Série 2 utilizando a 2ª estratégia.

A Fig. 14 apresenta o período de previsão para a Série 3, considerando a primeira estratégia. O período do histórico de previsão para esta série compreende o intervalo entre 31/12/2010 e 31/12/2011.

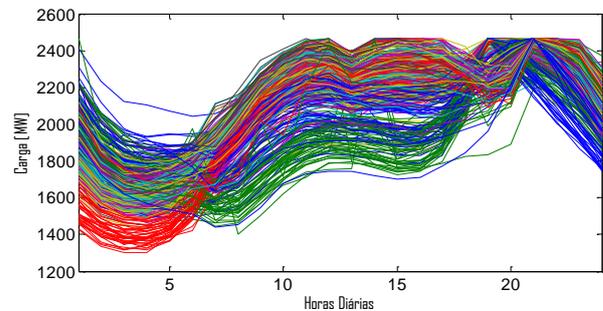


Figura 14 - Período de previsão para a Série 3 utilizando a 1ª estratégia.

A Fig. 15 apresenta este mesmo período, porém considerando a segunda estratégia.

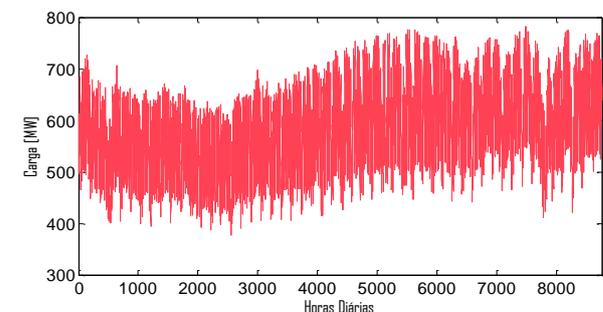


Figura 15 - Período de previsão para a Série 3 utilizando a 2ª estratégia.

Analisando os gráficos das séries percebe-se que visualmente a Série 1 é a que possui os dados mais comportados, enquanto que a Série 3 é a que possui a maior variação. Esta análise pode ser comprovada pelas informações contidas na Tab. 3, onde o coeficiente de variação do período de previsão da Série 1 é de 0,012 e o da Série 3 é 0,029.

## V. ESTUDO DE CASOS

Para testar a eficiência das estratégias propostas foram feitas simulações em diferentes períodos das 3 séries trabalhadas. A utilização de diferentes cenários visa a testar o comportamento dos previsores em diferentes situações, simulando o que ocorre no dia-a-dia dos agentes envolvidos no mercado de energia. Para cada período, o objetivo foi prever a demanda de carga para todo o ano e analisar a influencia da estratégia de previsão sobre o erro de previsão.

A previsão de demanda de carga elétrica não é uma tarefa trivial, pois possui uma grande quantidade de variáveis envolvidas, como as interferências exógenas já citadas, entre outros fatores, que dificultam a previsão de valores em longo prazo.

As estratégias de previsão foram testadas em períodos distintos das 3 séries. O período de previsão da Série 1 é de 31/10/2010 a 31/10/2011, da Série 2 foi de 31/03/2011 a 31/03/2012 e para a Série o período foi de 31/12/2010 a 31/12/2011.

Os resultados obtidos nos experimentos deste trabalho são apresentados nas subseções A e B.

**A. Resultados utilizando autocorrelação**

Esta subseção apresenta os resultados utilizando uma função de autocorrelação. Primeiramente são apresentados os resultados para a primeira estratégia e posteriormente para a segunda.

- Resultados utilizando a primeira estratégia

A Tab. 4 foi estruturada de modo que o erro percentual absoluto médio (MAPE) encontrado utilizando o método de previsão 1 passo a frente e o MAPE utilizando o método até n passos a frente são apresentados de acordo com a série estudada.

Tabela 4 – MAPE encontrados pela primeira estratégia utilizando uma função de autocorrelação.

MAPE	Séries		
	Série 1	Série 2	Série 3
1 passo	2,25%	2,35%	2,95%
n passos	2,89%	3,46%	5,20%

Analisando a Tab. 4 percebe-se que o método de previsão 1 passo a frente superou os resultados do método até n passos a frente para as 3 séries. O menor MAPE encontrado foi para a Série 1 (2,25%). Este resultado já era esperado, a Série 1 possui os dados mais comportados e o erro tende a ser maior quando se utiliza previsão sobre previsão (n passos), onde este é propagado a cada previsão.

As Fig. 16, 17 e 18 apresentam os valores reais e a melhor previsão para as Séries 1, 2 e 3, respectivamente, considerando a primeira estratégia.

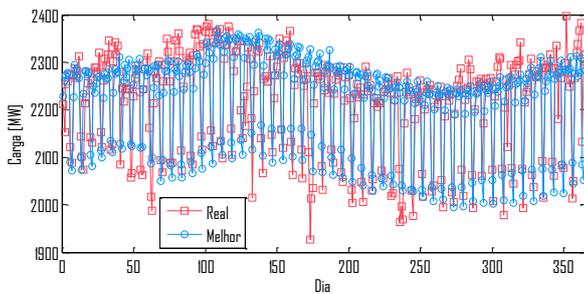


Figura 16 – Melhor resultado para a Série 1 utilizando autocorrelação.

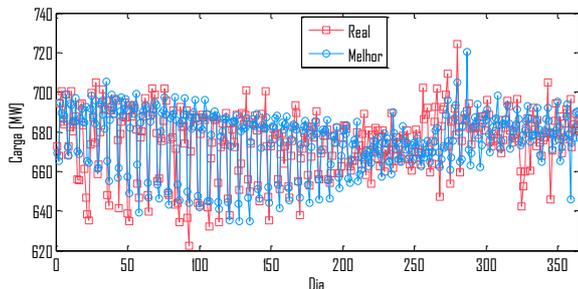


Figura 17 – Melhor resultado para a Série 2 utilizando autocorrelação.

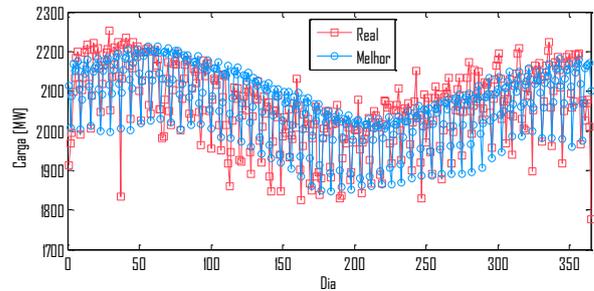


Figura 18 – Melhor resultado para a Série 3 utilizando autocorrelação.

- Resultados utilizando a segunda estratégia

A Tab. 5 foi estruturada de modo que o MAPE encontrado utilizando o método de previsão 1 passo a frente e o MAPE utilizando o método até n passos a frente são apresentados de acordo com a série estudada.

Tabela 5 - MAPE encontrados pela segunda estratégia utilizando uma função de autocorrelação.

MAPE	Séries		
	Série 1	Série 2	Série 3
1 passo	4,48%	4,45%	4,82%
n passos	11,76%	10,30%	12,60%

Analisando a Tab. 5 percebe-se que o método de previsão 1 passo a frente superou os resultados do método até n passos a frente para as 3 séries. O menor MAPE encontrado foi para a Série 2 (4,45%). Percebe-se também que nesta segunda estratégia, o método de previsão até n passos a frente apresentou um erro bem maior em relação a primeira estratégia para as três séries trabalhadas.

As Fig. 19, 20 e 21 apresentam os valores reais e a melhor previsão para as Séries 1, 2 e 3, respectivamente, considerando a segunda estratégia.

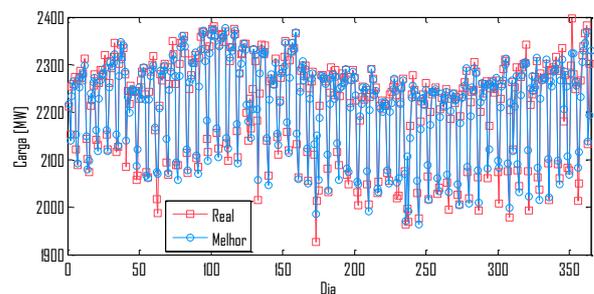


Figura 19 - Melhor resultado para a Série 1 utilizando autocorrelação.

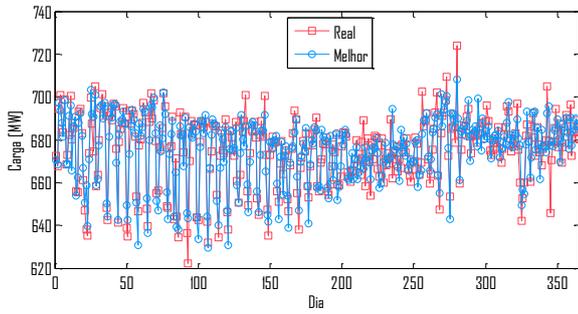


Figura 20 - Melhor resultado para a Série 2 utilizando autocorrelação.

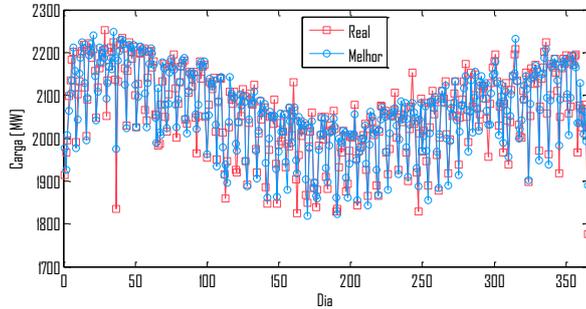


Figura 21 - Melhor resultado para a Série 3 utilizando autocorrelação.

**B. Resultado Utilizando 3, 5 e 7 atrasos**

Esta subsecção apresenta os resultados utilizando 3,5 e 7 atrasos. Primeiramente são apresentados os resultados para a primeira estratégia e posteriormente para a segunda.

- Resultados utilizando a primeira estratégia

A Tab. 6 foi estruturada de modo que o MAPE encontrado utilizando o método de previsão 1 passo a frente, o MAPE utilizando até o método até n passos a frente, variando o número de atrasos, são apresentados na primeira coluna. As demais colunas apresentam qual série foi utilizada.

Tabela 6 - MAPE encontrados pela primeira estratégia utilizando 3,5 e 7 atrasos

MAPE	Séries Tratadas		
	Série 1	Série 2	Série 3
1 passo e 3 atrasos	2,08%	2,29%	2,56%
n passos e 3 atrasos	3,02%	3,50%	4,79%
1 passo e 5 atrasos	2,11%	2,27%	2,64%
n passos e 5 atrasos	2,96%	3,51%	4,79%
1 passo e 7 atrasos	2,19%	2,31%	2,70%
n passos e 7 atrasos	2,90%	3,53%	4,79%

Analisando a Tab. 6 percebe-se que o método de previsão 1 passo a frente superou os resultados do método até n passos a frente para as 3 séries. O menor MAPE encontrado para as Séries 1e 3 foi utilizando 3 atrasos (2,08% e 2,56%), enquanto que para a Série 2, o menor MAPE foi utilizando 5 atrasos (2,27%).

As Fig. 22, 23 e 24 apresentam os valores reais e a melhor previsão para as Séries 1, 2 e 3, respectivamente, considerando a primeira estratégia.

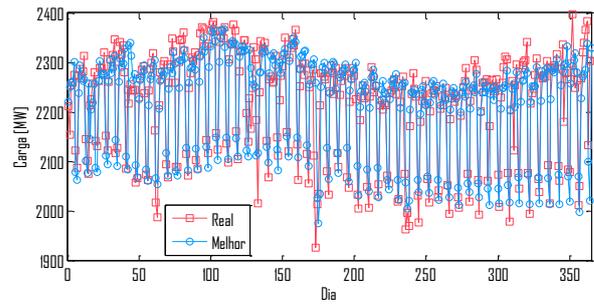


Figura 22 - Melhor resultado para a Série 1 utilizando atrasos.

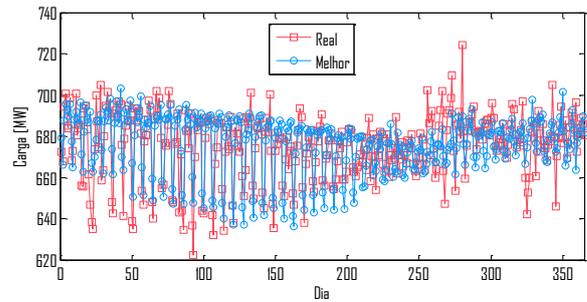


Figura 23 - Melhor resultado para a Série 2 utilizando atrasos.

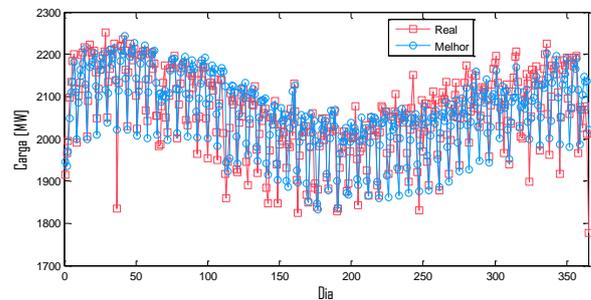


Figura 24 - Melhor resultado para a Série 3 utilizando atrasos.

- Resultados utilizando a segunda estratégia

A Tab. 7 foi estruturada de modo que o MAPE encontrado utilizando o método de previsão 1 passo a frente, o MAPE utilizando até o método até n passos a frente, variando o número de atrasos, são apresentados na primeira coluna. As demais colunas apresentam qual série foi utilizada.

Tabela 7 - MAPE encontrados pela segunda estratégia utilizando 3,5 e 7 atrasos

MAPE	Séries Tratadas		
	Série 1	Série 2	Série 3
1 passo e 3 atrasos	2,78%	2,73%	2,86%
n passos e 3 atrasos	11,76%	10,31%	12,59%
1 passo e 5 atrasos	2,71%	2,98%	2,92%
n passos e 5 atrasos	11,75%	10,30%	12,58%
1 passo e 7 atrasos	2,97%	2,82%	3,07%
n passos e 7 atrasos	11,75%	10,30%	12,59%

Analisando a Tab. 7 percebe-se que, assim como ocorreu na primeira estratégia, o método de previsão 1 passo a frente superou os resultados do método até n passos a frente para as 3 séries. O menor MAPE encontrado para as Série 1 foi utilizando 5 atrasos (2,71%), enquanto que para as Séries 2 e 3, o menor MAPE foi utilizando 3 atrasos (2,73% e 2,86%).

As Fig. 25, 26 e 27 apresentam os valores reais e a melhor previsão para as Séries 1, 2 e 3, respectivamente, considerando a segunda estratégia.

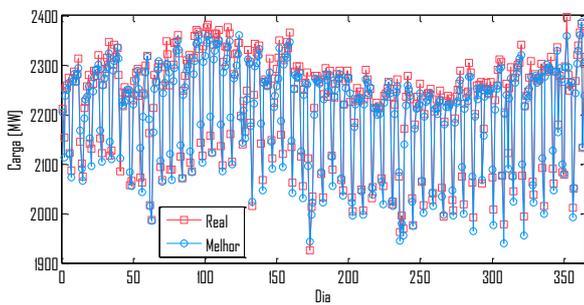


Figura 25 - Melhor resultado para a Série 1 utilizando atrasos.

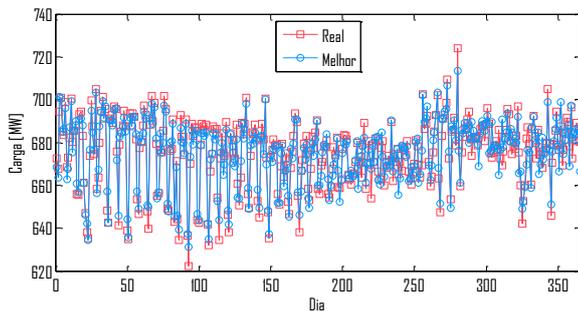


Figura 26 - Melhor resultado para a Série 2 utilizando atrasos.

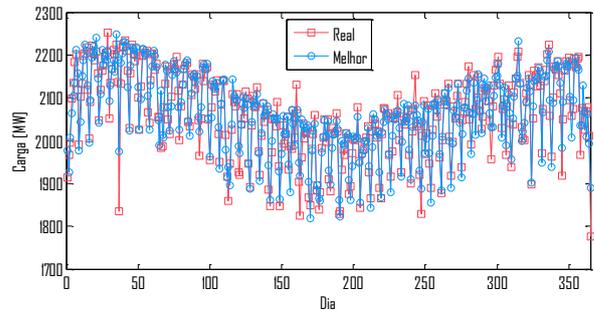


Figura 27 – Melhor resultado para a Série 3 utilizando atrasos.

Neste contexto, percebe-se que utilizar uma função de autocorrelação para escolher as entradas de cada saída desejada não melhorou os resultados para as séries trabalhadas.

A Série 1 foi a série que apresentou os melhores resultados para 3 dos 4 casos apresentados e a Série 3 apresentou os piores resultados para todos os casos. Este resultado já era esperado, pois na análise de dados, foi possível identificar que os dados mais comportados são da primeira série e os menos comportados da terceira série.

## VI. CONCLUSÃO

Este trabalho apresentou a construção de duas estratégias de previsão de carga elétrica num período de 1 ano a frente. Este tema é relevante no sentido de colaborar para o planejamento do sistema elétrico balanceando os custos das concessionárias e o atendimento das necessidades da demanda.

As análises realizadas demonstraram que os menores erros de previsão foram atribuídos aos dados mais comportados e os piores aos dados que possuíram as maiores variações.

De um modo geral, o método de previsão 1 passo a frente apresentou melhores resultados para as séries trabalhadas. A primeira estratégia se adaptou melhor aos dois métodos de previsão, apresentando bons resultados tanto para o método de previsão 1 passo a frente quanto para o método até n passos a frente.

A utilização da função de autocorrelação não se mostrou útil na diminuição do erro de previsão para as séries trabalhadas. Considerando os melhores resultados de cada experimento, as estratégias utilizando a função de autocorrelação não superaram, para nenhuma das séries, os melhores resultados encontrados utilizando atrasos.

Considerando os resultados apresentados, a estratégia mais adequada para as séries trabalhadas é a primeira, onde a previsão é realizada a cada hora individualmente. O melhor método de previsão a ser utilizado é o 1 passo a frente escolhendo as entradas de cada saída desejada através de atrasos.

## REFERÊNCIAS

- [1] ANEEL. Manual dos Programas de Pesquisa e Desenvolvimento Tecnológico do Setor Elétrico Brasileiro, Agência Nacional de Energia Elétrica, 2001.
- [2] A. J. Passini, Guide to Electrical Power Distribution System, CRC, 6<sup>th</sup> edition, 2005.

- [3] V. Iyer, C. C. Fung, T. A. Gedeon, A Fuzzy Neural Approach to Electricity Load and Spot-Price Forecasting in a Deregulated Electricity Market, Tencon, Conference on Converging Technologies for Asia-Pacific Region, vol. 4, pp. 1497-1482, October 2003.
- [4] Y. Guo, D. Niu, Y. Chen, Support Vector Machine model in Electricity Load Forecasting, Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, August, 2006.
- [5] D. Niu, Q. Wang, J. Li, Short Term Load Forecasting Model Using Support Vector Machine Based on Artificial Neural Network, Proceeding of the Fourth International Conference on Machine Learning and Cybernetics, vol. 7, pp. 4260-4265, August, 2005.
- [6] S. Santoso, H. W. Beaty, R. C. Dugan, M. M. McGranaghan, Electrical power System Quality, McGraw-Hill Professional, 2 ed., 2002.
- [7] M. H. Choueiki, C. A. Mount-Campbel, S. C. Ahalt, Building a 'Quasi Optimal' Neural Network to solve the Short-Term Load Forecasting Problem, IEEE Transaction on Power System, 12 (4), pp. 1432-1439, 1997.
- [8] A. G. Barkirtzis, V. Petridis, S. J. Alexiadis, A. H. Maissis, A Neural Network Short Term Load Forecasting Model for the Greek Power System, IEEE Transactions on Power Systems, 11 (2), pp. 858-863, 1995.
- [9] B. Xiao, P. G. McLaren, An Artificial Neural Network for Short Term Load Forecasting, IEEE/WESCANEX'95 Proceedings, pp. 129-132, 1995.
- [10] R. S. Zebulum, M. A. Pacheco, M. Vellasco, Um Sistema Inteligente de Previsão de Carga Aplicado ao Setor Elétrico Brasileiro, V Simpósio de Especialistas em Planejamento da Operação e Expansão Elétrica, Recife, pp. 791-797, 1996.
- [11] A. B. Altran, M. L. M. Lopes, C. R., F. Villarreal, Um Estudo de Funções de Base Radial Aplicadas à Previsão de Cargas Elétricas via Redes Neurais Artificiais, Congresso Nacional de Matemática Aplicada e Computacional, XVIII Cnmac, São Paulo – SP, 2005.
- [12] R. M. Salgado, Sistema de Suporte à Decisão para Previsão de Carga por Barramento. Tese de doutorado, Unicamp, 2009.
- [13] M. A. Cicogna, Sistema de Suporte à Decisão para Planejamento e Programação da Operação de Sistemas de Energia Elétrica, PhD thesis, FEEC/Unicamp, 2003.
- [14] R. M. Salgado, J. J. F. Pereira, T. Ohishi, R. Ballini, C. A. M. Lima, F. J. Von Zuben, A Hybrid Ensemble Model Applied to the Short-Term Load Forecasting Problem, Wcci-Ieee World Congress On Computational Intelligence, Vancouver, Canadá, 2006.
- [15] M. A. Leone Filho, Previsão de carga de curto prazo usando ensemble de previsores selecionados e evoluídos por algoritmos genéticos, dissertação de mestrado, Unicamp, 2006.
- [16] J. A. Gomes et al., Previsão de carga utilizando Support Vector Machine (SVM), AES Eletropaulo metropolitana, Mackenzie, Cidel, Argentina, 2010.
- [17] A. B. Altran, Sistema Inteligente para Previsão de Carga Multinodal em Sistemas Elétricos de Potência, Tese de doutorado, Unesp – Ilha Solteira, 2010.
- [18] K. Siwek, S. Osowski, Short Term Load Forecasting Model In The Power System Using Ensemble Of Predictors, Instrumentation and Measurement Technology Conference – Imtc 2007 Warsaw, Poland, May 1-3, 2007.
- [19] I. Drezga, S. Rahman, Short-Term Load Forecasting, IEEE Transaction On Power System, vol. 14, n. 3, August 2009.
- [20] M. E. Telbany, F. E. Karmi, Short-Term Forecasting of Jordanian Electricity Demand Using Particle Swarm Optimization, Electric Power System Research, Lausanne, v.78, n.3, pp. 425-433, 2008.
- [21] Z. Yun, Z. Quan, S. Caixin, L. Shaolan, L. Yuming, S. Yang, RBF Neural Network and ANFIS-based Short-Term Load Forecasting Approach in Real-Time Price Environment, IEEE Transaction on Power System, New York, v.23, n.3, pp. 853-858, 2008.
- [22] V. H. Ferreira, A. P. A. da Silva, Modelos Baseados em Kernel para Previsão de Carga a Curto Prazo, IX Congresso Brasileiro de Redes Neurais/Inteligência Computacional, Ouro preto, 25-28 de Outubro, 2009.
- [23] R. J. Campos, Previsão de Série Temporais com Aplicações a Séries de Consumo de Energia Elétrica, Dissertação de Mestrado, UFMG, Belo Horizonte, abril de 2008.
- [24] R. M. Salgado, Um modelo de Previsão de Carga por Barramento, Dissertação de Mestrado, Unicamp, 2004.
- [25] M. O. Moreira, Sistema Inteligente Baseado em Combinação de Previsores para Previsão de Séries Temporais de Carga Elétrica à Curto Prazo.
- [26] A. J. R. Reis, A. P. A. da Silva, Aplicação da Transformada Wavelet Discreta na Previsão de Carga à Curto Prazo Via Redes Neurais, Revista Controle e Automação, v. 15, n. 1, abril de 2003.
- [27] T. B. S. Everit, The Cambridge dictionary of statistics in the medical sciences. Cambridge: Cambridge University Press, 1995.
- [28] P. A. Morettin, C. M. C. Toloi, Previsão de Série Temporais, 2 ed., São Paulo, Atual Editora, 1985.
- [29] T. M. Mitchell, Machine Learning, New York, McGraw-Hill, 1997.
- [30] S. R. Gunn, M. Brown, K. M. Bossley, Network performance assessment for neurofuzzy data modeling advanced in intelligent data analysis, vol. 1280, pp. 313-323, 1997.
- [31] A. C. Lorena, A. C. P. L. F., Uma Introdução às Support Vector Machines, Revista de Informática Teórica e Aplicada, vol. 14, n. 2, 2007.
- [32] D. de Braga, M. D. C. Lima, R. M. Salgado, An empirical analysis of mlp neural network applied to streamflow forecasting, Latin America transaction, IEEE 9 (3), pp. 295-301, 2011.
- [33] V. Vapnik, V. Lerner, Pattern Recognition Using Generalized Portrait Method, Automation and Remote Control, pp. 774-778, 1963.
- [34] V. Vapnik, A. Chervonenkis, Statistical Learning Theory, Wiley – Interscience, 1964.
- [35] R. Semolini, Support Vector Machine, Inferência Transdutiva e o Problema de Classificação, Tese de Doutorado, FEEC, Unicamp, 2002.
- [36] V. Vapnik, A. Chervonenkis, Statistical Learning Theory, Wiley-Interscience, 1998.
- [37] V. Vladimir, V. Vapnik, The Nature of Statistical Learning Theory, 1994.
- [38] A. J. Smola, B. Schölkopf, Sparse greedy matrix approximation for machine learning, Citeseer, 2000.
- [39] G. C. A. Coutinho, Reconhecimento Facial Baseado em Support Vector Machines, Trabalho de Conclusão de Curso, UFPA, 2007.
- [40] A. R. G. Lima, Máquinas de Vetores Suporte na Classificação de Impressões Digitais, Dissertação de Mestrado, Universidade Federal do Ceará, 2002.
- [41] F. Parrela, Online Support Vector Regression, Dissertação de Mestrado, University of Genoa, 2007.
- [42] C. Cortes, V. Vapnik, Support Vector Networks, Machine Learning, pp. 273-297, 1995.
- [43] H. F. Ferro, Otimização da Previsão de Carga Elétrica a Curto Prazo Utilizando Critérios de Similaridade entre Perfis de Consumo, Dissertação de Mestrado, Universidade Federal de Santa Catarina, 2007.
- [44] A. J. Smola, B. Schölkopf, A Tutorial on Support Vector Regression, Statistic and Computing, pp. 199-222, 2002.
- [45] H. Wei, N. Yoshiteru, W. Shouyang, A general approach based on autocorrelation to determine input variables os neural networks for time series forecasting, Journal of System Science and Complexity, v.17, n. 3, July, 2004.
- [46] V. L. do Forte, G. P. de Melo, R. Mariano, Aplicação da Função de Autocorrelação na Determinação de Parâmetros de Entrada em Rede RBF na Predição de Séries Temporais. Não publicado.
- [47] C. C. Chang, C. J. Lin, LIBSVM: a Library for Support Vector Machines, ACM Transactions on Intelligent System and Technology 2, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [48] R. Fan, P. Chen, C. Lin, Working set selection using second order information for training support vector machines, The Journal Machine Learning Research 6: pp. 1889-1918, 2005.