

## 2 – REVISÃO BIBLIOGRÁFICA

### 2.1 – Introdução

Nesta parte do trabalho, serão apresentados alguns conceitos de séries temporais, assim como alguns modelos estatísticos e modelos de inteligência computacional que são utilizados para previsão. Além disso, será apontado como estes modelos podem ser estendidos para dados de alta frequência.

#### 2.1.1 – Série Temporal

Uma série temporal é uma coleção de observações de uma certa variável aleatória, ordenadas seqüencialmente no tempo, em intervalos equidistantes (a cada minuto, a cada quinze minutos, horário, diário, semanal, mensal, trimestral, anual), de tal forma que exista dependência serial ao longo do tempo. Denominando-se  $Z_t$  o valor da variável aleatória no instante  $t$ , a série temporal pode ser escrita por  $Z_1, Z_2, \dots, Z_T$ , sendo  $T$  o tamanho da série, o qual reflete o número de observações da variável (Souza e Camargo, 2004).

Este conceito de série temporal é bem distinto do de amostra aleatória. Nesse último, tem-se uma população e cada elemento dessa tem uma probabilidade (diferente de zero) de ser incluída na amostra (subconjunto da população). Por exemplo, o médico, ao pesar e medir uma determinada criança todos os meses, ao longo de um ano, terá construído uma série temporal de tamanho 12. Em contra partida, o ato de selecionar (através de técnicas de amostragem) 12 crianças de uma escolinha, a fim de verificar qual é a altura e o peso médio das crianças dessa escola, caracteriza uma amostra aleatória.

Em séries temporais, podem-se usar duas perspectivas:

- Análise no Domínio do Tempo

Neste caso, é considerada a evolução temporal do processo, isto é, o interesse está voltado para a dimensão do evento no instante de tempo. Para este tipo de avaliação, são usadas as funções de autocovariância e autocorrelação.

A autocovariância é a covariância da série com ela mesma, porém em um intervalo defasado em  $k$ , podendo ser escrito da seguinte forma: covariância de  $Z_t$  com  $Z_{t+k}$ :

$$\gamma_k = Cov[Z_t, Z_{t+k}] = E[(Z_t - \mu)(Z_{t+k} - \mu)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (Z_t - \mu)(Z_{t+k} - \mu)P(Z_t, Z_{t+k})dZ_t dZ_{t+k}$$

Sendo:  $\mu$  a média do processo e  $P(Z_t, Z_{t+k})$  a função de densidade de probabilidade conjunta das variáveis  $Z_t$  e  $Z_{t+k}$ .

Para uma série temporal, como já definida, a autocovariância pode ser estimada por:

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})$$

Sendo  $\bar{Z}$  a média de  $Z_t$  descrita por:

$$\bar{Z} = \frac{1}{T} \sum_{t=1}^T Z_t$$

A autocorrelação mostra a quantidade de defasagens que dependem de  $t$  e é definida como sendo a autocovariância padronizada:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov[Z_t, Z_{t+k}]}{\sqrt{Var(Z_t)Var(Z_{t+k})}}$$

Sendo:

$$\gamma_0 \text{ a variância e } Var(Z_t) = Var(Z_{t+k})$$

$$\rho_0 = 1$$

$$\rho_k = \rho_{-k}$$

Pode-se estimar a autocorrelação de uma série da seguinte forma:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{T-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^T (Z_t - \bar{Z})^2}$$

Há também a autocorrelação parcial, que analisa a correlação serial apenas entre  $Z_t$  e  $Z_{t+k}$ , não considerando as observações intermediárias, podendo, então, ser representada da seguinte forma:

$$Cor(Z_t, Z_{t+k} | Z_{t+1}, \dots, Z_{t+k-1})$$

- **Análise no Domínio da Frequência**

Neste caso, a análise é feita sobre a frequência com que os eventos ocorrem e é utilizada quando os componentes harmônicos da série são mais significantes. Para tal, é utilizada a densidade espectral, com a qual é possível extrair as propriedades relacionadas à frequência.

Uma vez que o espectro de um processo não é conhecido, é necessário estimá-lo, o que é, geralmente, feito através do periodograma de janelas espectrais.

### 2.1.2 – Processo Estocástico

Uma série temporal é uma realização de um processo estocástico. Assim, uma série temporal  $y_t$  pode ser definida como uma função  $y$  da variável independente  $t$ , que é gerada por um processo estocástico desconhecido (Souza e Camargo, 2004).

Um processo estocástico é definido como sendo um conjunto de variáveis aleatórias definidas por uma lei de probabilidades.

Como o processo estocástico é um mecanismo para gerar dados que não é descrito por uma função determinística, seu comportamento futuro tem que ser descrito probabilisticamente. O processo somente pode ser estatisticamente determinado quando todas as suas funções de distribuição de probabilidade são conhecidas até a T-ésima ordem (Barros, 2004).

Na prática, não se conhecem todas as funções de probabilidade até a T-ésima ordem e, em regra, tem-se somente uma realização do processo estocástico, a partir do qual se deseja inferir sobre o mecanismo gerador da série. Devido a isto, assumem-se duas restrições para superar esses empecilhos: estacionariedade e ergodicidade.

No processo estocástico estacionário algumas características permanecem constantes ao longo do tempo, sendo que a estacionariedade pode ser do tipo forte ou fraca.

Na estacionariedade forte ou estrita a forma da distribuição conjunta do processo permanece sem variação mediante uma translação temporal. Como na prática é muito difícil especificar a distribuição conjunta de um processo estocástico, há a versão mais fraca, na qual somente alguns momentos do processo permanecem inalterados no tempo (Medeiros, 2005).

Assim, a estacionariedade fraca ou de segunda ordem ocorre quando a média e a variância do processo são constantes no tempo e sua estrutura de dependência linear depende apenas da distância entre os períodos (da defasagem), isto é:

$$\begin{aligned} E[Z_t] &= \mu & \forall t \\ \text{Var}[Z_t] &= \sigma^2 & \forall t \\ E[(Z_t - \mu)(Z_{t-k} - \mu)] &= \gamma_k & \forall t \end{aligned}$$

A ergodicidade ocorre quando uma única realização de um processo estocástico é suficiente para obter todas as suas estatísticas. Por isso, todo processo ergótico é também estacionário, visto que se for não estacionário, não conterá todas as informações necessárias para a especificação do processo.

A série temporal e o processo estocástico possuem uma relação análoga à de uma amostra e da população da qual ela foi extraída. Então, o objetivo da análise de uma série temporal é retirar de uma realidade (processo estocástico) uma amostra finita de dados equidistantes no tempo (série temporal) e identificar um modelo que seja capaz de inferir sobre o comportamento da realidade (processo estocástico).

### 2.1.3 – Previsão de Séries Temporais

O modelo ajustado pode ser usado para fazer previsões dos valores futuros. Assim, baseados nas informações passadas e atuais, são estimados, de forma quantitativa, os eventos futuros.

Em uma previsão, o horizonte é o intervalo de tempo à frente, a partir da última observação, para o qual os valores são estimados. Além disso, denomina-se

número de passos à frente da previsão o número de intervalos previstos a partir da origem (Souza, 1989).

Dessa forma, a previsão denotada por  $\hat{Z}_T(k)$  é definida como a esperança condicional de  $Z_{T+k}$  dados os valores passados:

$$\hat{Z}_T(k) = E(Z_{T+k} | Z_T, Z_{T-1}, \dots)$$

Sendo  $Z_{T+k}$  os valores a serem previstos para  $k=1,2,\dots$

Uma previsão quantitativa pode ser caracterizada através da sua origem, do número de passos à frente, de seu valor pontual e de uma medida de incerteza a ela associada.

Para definir os valores futuros da série em questão, é necessário formular um modelo matemático que seja capaz de representar o comportamento e as características da série em questão. Há várias técnicas que podem ser empregadas na previsão de séries temporais, mais especificamente na previsão de carga elétrica, objeto desse estudo. A seguir, serão apresentadas algumas dessas técnicas.

Têm-se os modelos univariados que utilizam somente os valores passados da própria série para explicar os valores futuros. Dentro desses, há os métodos de decomposição, métodos de amortecimento exponencial e os modelos de Box & Jenkins. Um exemplo é utilizar somente os valores passados de carga de energia elétrica para prever a própria carga.

Já os modelos causais, ou modelos de função de transferência, usam os valores passados da própria série a ser prevista e de outras séries que possuem relação com o que se quer prever. Neste caso, poder-se-ia prever a carga de energia elétrica usando seu histórico e a série de temperatura, por exemplo.

Há também os modelos multivariados, com os quais é possível fazer previsões de várias séries ao mesmo tempo como, por exemplo, prever a carga de energia de várias concessionárias simultaneamente.

Os modelos de inteligência artificial também podem ser usados em previsão. O objetivo desses modelos é desenvolver algoritmos que sejam capazes de reproduzir tarefas que são realizadas por seres humanos e necessitam de cognição, como o raciocínio, a aprendizagem e o auto-aperfeiçoamento (Haykin, 1999).

## 2.2 – Método Ingênuo (NAÍVE)

Este método, também conhecido como *random walk* (passeio aleatório), é o mais simples para realizar a previsão de uma variável, no qual o valor previsto será igual ao último valor observado (independente de quantos passos à frente). Portanto, este método possui a seguinte equação:

$$Z_t = Z_{t-1} + \varepsilon_t$$

Denotando a previsão  $\tau$  passos à frente a partir do instante T por  $\hat{Z}_T(\tau)$ , ou seja, o valor previsto para  $Z_{T+\tau}$ , pode-se escrever a equação de previsão da seguinte forma:

$$\hat{Z}_T(\tau) = E[(Z_{T+\tau} + \varepsilon_{T+\tau}) | Z_T] = E(Z_{T+\tau})$$

Por sua simplicidade, o método ingênuo é usado como critério de comparação (*benchmark*), isto é, o modelo a ser testado tem que ser sempre melhor do que este. Para tal, usa-se a estatística GMRAE, *Geometric Mean of the Relative Absolute Error* (Armstrong & Collopy, 1992).

$$GMRAE_{m,h} = \left[ \prod_{s=1}^N RAE_{m,h,s} \right]^{1/N}$$

$$RAE_{m,h,s} = \frac{|F_{m,h,s} - A_{h,s}|}{|F_{rw,h,s} - A_{h,s}|}$$

Sendo: m = o método de previsão;

rw = o método ingênuo;

h = o horizonte de previsão;

s = a série a ser prevista;

$F_{m,h,s}$  = é a previsão feita pelo método m, para a série s, h passos à frente;

$A_{h,s}$  = é o valor da série s, h passos à frente;

N = é o número de séries utilizadas;

RAE = erro absoluto relativo (*relative absolute error*).

Para dados de alta frequência com sazonalidade, o método ingênuo possui o seguinte modelo para previsão:

$$\hat{Z}_T(\tau) = E(Z_{T+\tau-S})$$

Sendo S a sazonalidade.

### 2.3 – Média Móvel

No método de média móvel, cada nova observação é calculada pela média de  $n$  ( $n$  é denominado o tamanho da janela) pontos anteriores (Montgomery, 1990):

$$Z_t = \frac{1}{n}(Z_{t-1} + Z_{t-2} + \dots + Z_{t-n})$$

Neste caso, o tamanho da janela é o parâmetro a ser ajustado, sendo necessário buscar um equilíbrio, pois quanto maior o tamanho da janela, mais suave será o comportamento da média e, conseqüentemente, mais imune a ruídos e movimentos curtos ela estará. Porém, se o tamanho da janela for muito grande, a série prevista irá reagir de maneira muito lenta às mudanças significativas.

Com base nisso, pode-se ter um modelo de média móvel simples, no qual a série é caracterizada localmente por seu nível, sendo as mudanças ocorridas, nesse tipo de caso, de caráter muito lento. Há a média móvel dupla, onde a série possui tendência aditiva, sendo que a média dessa série sofre uma alteração linear ao longo do tempo. Já a média móvel tripla, é utilizada em séries com tendência quadrática, sendo seu efeito multiplicativo.

O modelo de média móvel simples é:

$$Z_t = a + \varepsilon_t$$

E a equação de previsão  $\tau$  passos à frente, feita no instante  $T$  é:

$$\hat{Z}_T(\tau) = \hat{a}(T)$$

Nele, a série oscila em torno de um nível médio constante, nesse caso “ $a$ ”, o qual precisa ser estimado. Considerando  $\hat{a}(T)$  o estimador do parâmetro “ $a$ ” no instante  $T$ , tem-se:

$$\hat{a}(T) = M_T = \frac{Z_T + Z_{T-1} + \dots + Z_{T-N+1}}{N}$$

O modelo para a média móvel dupla é:

$$Z_t = a_1 + a_2 t + \varepsilon_t$$

Sendo o modelo de previsão  $\tau$  passos à frente feita em  $T$  igual a:

$$\hat{Z}_T(\tau) = \hat{a}_1(T) + \hat{a}_2(T)[T + \tau]$$

Os estimadores dos parâmetros no instante  $T$  são calculados por:

$$\hat{a}_1(T) = 2M_T - M_T^{[2]} - \hat{a}_2(T)T$$

$$\hat{a}_2(T) = \frac{2}{N-1} [M_T - M_T^{[2]}]$$

Sendo:

$$M_T = \frac{Z_T + Z_{T-1} + \dots + Z_{T-N+1}}{N} \quad (\text{m\u00e9dia m\u00f3vel simples de tamanho } N)$$

$$M_T^{[2]} = \frac{M_T + M_{T-1} + \dots + M_{T-N+1}}{N} \quad (\text{m\u00e9dia m\u00f3vel dupla de tamanho } N)$$

O modelo para a m\u00e9dia m\u00f3vel tripla \u00e9:

$$Z_t = a_1 + a_2t + a_3t^2 + \varepsilon_t$$

E a previs\u00e3o para este modelo,  $\tau$  passos \u00e0 frente, \u00e9:

$$\hat{Z}_T(\tau) = \hat{a}_1(T) + \hat{a}_2(T)[T + \tau] + \hat{a}_3(T)[T + \tau]^2$$

Sendo,  $\hat{a}_1(T), \hat{a}_2(T), \hat{a}_3(T)$  os estimadores dos par\u00e2metros no instante T, calculados pelas m\u00e9dias m\u00f3veis de tamanho N: simples ( $M_T$ ), dupla ( $M_T^{[2]}$ ) e tripla ( $M_T^{[3]}$ ).

O modelo de previs\u00e3o para dados com sazonalidade utilizando m\u00e9dias m\u00f3veis pode ser escrito por:

$$\hat{Z}_T(\tau) = \frac{1}{n} E(Z_{T+\tau-S} + Z_{T+\tau-2S} + \dots + Z_{T+\tau-nS})$$

## 2.4 – M\u00e9todos de Amortecimento Exponencial

Gardner (2006) traz em seu artigo uma descri\u00e7\u00e3o sobre o estado da arte dos m\u00e9todos de amortecimento exponencial at\u00e9 a atualidade. Neste trabalho, o autor relata que o m\u00e9todo de amortecimento exponencial foi inicialmente desenvolvido por Robert G. Brown no per\u00edodo em que ele trabalhava para a marinha norte-americana durante a II Guerra Mundial, onde utilizou este modelo para prever localidades de submarinos, com o intuito de intercept\u00e1-los.

Na d\u00e9cada de 50, Brown expandiu o modelo, desenvolvendo m\u00e9todos para a tend\u00eancia e a sazonalidade. Ainda nessa d\u00e9cada, e trabalhando independentemente de Brown, Charles C. Holt desenvolveu um m\u00e9todo similar ao

amortecimento exponencial considerando uma tendência aditiva (tal trabalho foi publicado há poucos anos em Holt (2004a) e Holt (2004b)). Contudo, essa idéia foi testada por Peter R. Winters (Winters, 1960) e, por isso, ficou denominada como método de Holt-Winters.

Os métodos de amortecimento exponencial são classificados como sendo automáticos e de validade local. Eles se baseiam na idéia de que as observações mais recentes contêm mais informações do que as observações mais antigas e, por isso, o peso decresce à medida que a observação torna-se antiga (a taxa de decréscimo é determinada por uma ou mais constantes de amortecimento), diferentemente do método das médias móveis em que todas as observações têm o mesmo peso, limitando o modelo (Souza, 1983).

Mesmo estando baseada em uma filosofia simples, essa classe de métodos produz, geralmente, bons resultados e, por isso, é frequentemente utilizada como referência no desenvolvimento de modelos de previsão.

Além disso, a robustez e acuracidade desses métodos têm aumentado suas formas de aplicação e utilização, principalmente em séries temporais que precisam ser atualizadas através de um procedimento automático.

#### 2.4.1 – Métodos de Amortecimento de Brown

Nesta classe, estão os modelos automáticos para séries não sazonais, podendo ser os modelos: simples, duplo e triplo. A diferença desses modelos para os de médias móveis está na estimação das constantes de amortecimento. Por isso, a seguir, será mostrado apenas como as constantes são estimadas em cada caso.

(i) Simples:

$$\hat{a}(T) = \alpha Z_T + (1 - \alpha)Z_{T-1} = \alpha Z_T + (1 - \alpha)\hat{a}(T - 1)$$

(ii) Duplo:

$$\begin{aligned}\hat{a}_1(T) &= 2S_T - S_T^{[2]} \\ \hat{a}_2(T) &= \frac{\alpha}{1 - \alpha} (S_T - S_T^{[2]}) \\ S_T &= \alpha Z_T + (1 - \alpha)S_{T-1} \\ S_T^{[2]} &= \alpha S_T + (1 - \alpha)S_{T-1}^{[2]}\end{aligned}$$

(iii) Triplo:

$$\hat{a}_1(T) = 3S_T - 3S_T^{[2]} + S_T^{[3]}$$

$$\hat{a}_2(T) = \left(\frac{\alpha}{1-\alpha}\right)^2 [S_T - 2S_T^{[2]} + S_T^{[3]}]$$

$$\hat{a}_3(T) = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)S_T - 2(5-4\alpha)S_T^{[2]} + (4-3\alpha)S_T^{[3]}]$$

$$S_T = \alpha Z_T + (1-\alpha)S_{T-1}$$

$$S_T^{[2]} = \alpha S_T + (1-\alpha)S_{T-1}^{[2]}$$

$$S_T^{[3]} = \alpha S_T^{[2]} + (1-\alpha)S_{T-1}^{[3]}$$

### 2.4.2 – Métodos de Amortecimento de Holt

Este método possui a mesma formulação que o anterior. O grande diferencial está na atualização dos parâmetros. Enquanto o primeiro usava o mesmo hiperparâmetro para atualizar cada parâmetro (no caso anterior, foi usado  $\alpha$  em todos), este método usa um hiperparâmetro diferente em cada parâmetro. Abaixo serão mostrados o modelo e as equações de atualização para o caso de uma série com tendência linear, sendo que o mesmo pode ser estendido para o caso de tendência quadrática (para tendência constante, os dois métodos são equivalentes dado que só terá um parâmetro a ser atualizado).

Modelo:

$$Z_t = a_1 + a_2 t + \varepsilon_t$$

Equação de previsão, considerando a origem deslocada para T:

$$\hat{Z}_T(\tau) = \hat{a}_1(T) + \hat{a}_2(T)\tau$$

Para esse caso, os estimadores de  $\hat{a}_1(T)$  e  $\hat{a}_2(T)$ , calculados via amortecimento exponencial com constantes  $\alpha$  e  $\beta$ , são:

$$\hat{a}_1(T) = \alpha Z_T + (1-\alpha)[\hat{a}_1(T-1) + \hat{a}_2(T-1)]$$

$$\hat{a}_2(T) = \beta[\hat{a}_1(T) - \hat{a}_1(T-1)] + (1-\beta)\hat{a}_2(T-1)$$

Ainda, pode-se ter a tendência amortecida, o que é conhecida como *damped trend* (Taylor, 2003b). Nesse caso, teria-se mais uma constante de amortecimento,  $\varphi$ , e a equação de previsão do modelo seria:

$$\hat{Z}_T(\tau) = \hat{a}_1(T) + \left[ \sum_{j=1}^{\tau} \varphi^{j-1} \right] \hat{a}_2(T)$$

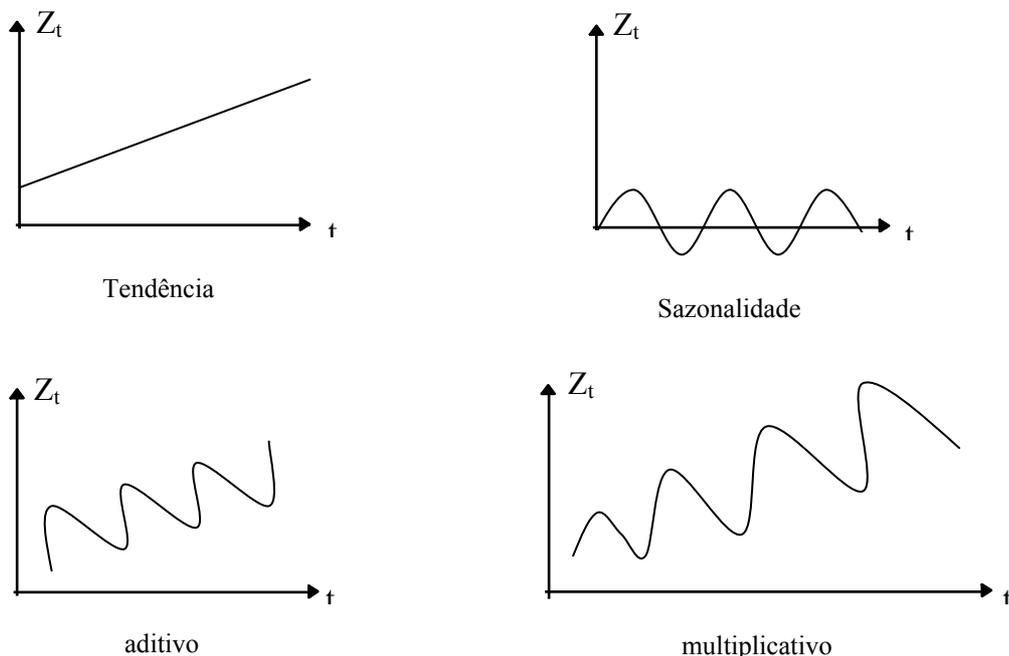
Assim, se  $0 < \varphi < 1$ , a tendência é amortecida (*damped*) e a previsão se aproxima da assíntota:  $a_1(t) + \frac{\varphi}{1-\varphi} a_2(t)$ . Caso contrário, se  $\varphi > 1$ , a tendência passa a ter um comportamento “exponencial”, que pode ser explosivo.

### 2.4.3 – Métodos de Amortecimento de Holt-Winters

Este método foi introduzido por Winters para modelar séries que apresentam variação cíclica.

Diferentemente dos dois métodos anteriores, que não incorporavam sazonalidade (eles tinham apenas tendência), este método inclui a sazonalidade em seu modelo, sendo que essa pode ser aditiva, que é recomendada para séries homocedásticas, ou multiplicativa, recomendada para séries heterocedásticas. Exemplos dessas componentes encontram-se na figura 2.01:

Figura 2.01 – Exemplo da componente tendência e sazonalidade



As séries homocedásticas são aquelas na qual a variância não cresce ao decorrer do tempo, sendo comum em séries físicas. As heterocedásticas possuem variância crescente, ocorrendo geralmente em séries de mercado.

## (i) Modelo Aditivo

Seja a série,  $Z_t$ , com tendência e um padrão sazonal. Pode-se escrever seu modelo da seguinte forma:

$$Z_t = a_1 + a_2 t + c_t + \varepsilon_t$$

Sendo,  $c_t$ , o fator sazonal.

Neste modelo, os fatores sazonais têm que seguir a seguinte restrição:

$$\sum_{i=1}^L c_i(t) = 0$$

Sendo:  $L$  o tamanho do fator sazonal.

Denotando  $\hat{a}_1(T), \hat{a}_2(T), \hat{c}_i(T)$  as estimativas dos componentes nível, tendência e fatores sazonais, respectivamente, no instante  $T$ , tem-se que para este instante, eles são calculados por:

$$\begin{aligned}\hat{a}_1(T) &= \alpha [Z_T - \hat{c}_{m(T)}(T-1)] + (1-\alpha)[\hat{a}_1(T-1) + \hat{a}_2(T-1)] \\ \hat{a}_2(T) &= \beta [\hat{a}_1(T) - \hat{a}_1(T-1)] + (1-\beta)\hat{a}_2(T-1) \\ \hat{c}_{m(T)}(T) &= \gamma [Z_T - \hat{a}_1(T)] + (1-\gamma)\hat{c}_{m(T)}(T-1)\end{aligned}$$

Sendo  $\alpha, \beta$  e  $\gamma$  as constantes de amortecimento.

Assim, a equação de previsão, para  $\tau$  períodos à frente, feita no instante  $T$ , pode ser escrita por:

$$\hat{Z}_T(\tau) = \hat{a}_1(T) + \hat{a}_2(T)\tau + \hat{c}_{m(T+\tau)}(T)$$

Para a utilização deste modelo, é necessário que os parâmetros iniciais  $\hat{a}_1(0), \hat{a}_2(0), \hat{c}_i(0)$ ,  $i=1, \dots, L$ , sejam pré-calculados. A estimativa desses parâmetros iniciais pode ser obtida através dos dados históricos e estão descritos de forma detalhada em Montgomery (1990).

## (ii) Modelo Multiplicativo

Esta é a versão do modelo para o caso multiplicativo, que tem a premissa que a amplitude da sazonalidade varia no tempo. Associado a isto, ela continua possuindo a formulação aditiva para a componente tendência, assim este modelo incorpora a tendência linear com o efeito sazonal multiplicativo.

Então, este modelo pode ser escrito por:

$$Z_t = (a_1 + a_2 t)c_t + \varepsilon_t$$

Aqui, os fatores sazonais são tais que a sua soma é igual ao comprimento da sazonalidade, isto é:

$$\sum_{i=1}^L c_i(t) = L$$

Então, denotando por  $\hat{a}_1(T), \hat{a}_2(T), \hat{c}_i(T)$ , para  $i=1, \dots, L$ , o nível, a tendência e os fatores sazonais, respectivamente, estimados em  $T$ , estes podem ser atualizados, admitindo o conhecimento de seus resultados em  $T-1$ , por:

$$\begin{aligned}\hat{a}_1(T) &= \alpha \left( \frac{Z_T}{\hat{c}_{m(T)}(T-1)} \right) + (1-\alpha)(\hat{a}_1(T-1) + \hat{a}_2(T-1)) \\ \hat{a}_2(T) &= \beta(\hat{a}_1(T) - \hat{a}_1(T-1)) + (1-\beta)\hat{a}_2(T-1) \\ \hat{c}_{m(T)}(T) &= \gamma \left( \frac{Z_T}{\hat{a}_1(T)} \right) + (1-\gamma)\hat{c}_{m(T)}(T-1)\end{aligned}$$

Sendo:  $\alpha, \beta$  e  $\gamma$  as constantes de amortecimento;

os demais fatores sazonais:  $\hat{c}_i(T) = \hat{c}_i(T-1)$ ;  $i=1, 2, \dots, L$ , para  $i \neq m(T)$ .

Assim como no modelo aditivo, é necessário estimar os parâmetros iniciais:  $\hat{a}_1(0), \hat{a}_2(0), \hat{c}_i(0)$ ,  $i=1, \dots, L$ , os quais estão detalhados em Montgomery (1990).

Neste trabalho será utilizada uma extensão do modelo de Holt-Winters em dados de alta frequência. Essa metodologia está detalhada no próximo capítulo.

## 2.5 – Método de Amortecimento Direto

O método de amortecimento direto é utilizado na estimação dos parâmetros que compõe os modelos definidos como combinações lineares de funções do tempo. Essa estimação é feita através dos mínimos quadrados ponderados Souza (1983).

Assumindo uma série temporal  $Z_t$ , o modelo pode ser representado por:

$$Z_t = \sum_{i=1}^k a_i f_i(t) + \varepsilon_t ; t = 1, 2, \dots, T$$

Onde:  $a_i$ :  $i = 1, 2, \dots, k$  são os parâmetros do modelo;

$f_i(T)$ :  $i = 1, 2, \dots, k$  são funções matemáticas do tempo;

$\varepsilon_t$ : ruído do sistema, sendo este independente e identicamente distribuído com média nula e variância constante  $\sigma_\varepsilon^2$ ;

$T \geq k$ .

Pode-se reescrever a fórmula acima da seguinte forma:

$$Z_t = f^T(t)a + \varepsilon_t$$

Sendo:  $a = [a_1 a_2 \dots a_k]^T$ ;

$$f(t) = [f_1(t) f_2(t) \dots f_k(t)]^T$$

$[\cdot]^T$  significa o transposto de  $[\cdot]$ .

O estimador de mínimos quadrados ponderados (MQP) é definido pela minimização da seguinte função:

$$SSE = \sum_{t=1}^T w_t^2 \varepsilon_t^2$$

As quantidades  $Z$ ,  $f(T)$ ,  $\varepsilon$  e  $w$  podem ser reescritas da seguinte forma:

$$Z_{(T \times 1)} = [Z_1 Z_2 \dots Z_T]^T$$

$$\varepsilon_{(T \times 1)} = [\varepsilon_1 \varepsilon_2 \dots \varepsilon_T]^T$$

$$f(T)_{(T \times T)} = \begin{bmatrix} f_1(1) f_2(1) \Lambda f_k(1) \\ f_1(2) f_2(2) \Lambda f_k(2) \\ \text{M} \\ f_1(T) f_2(T) \Lambda f_k(T) \end{bmatrix}$$

$$w_{(T \times T)} = \begin{bmatrix} w_{11} & & & 0 \\ & w_{22} & & \\ & & \ddots & \\ 0 & & & w_{TT} \end{bmatrix}$$

Assim, é possível reescrever a equação de minimização do erro da seguinte forma:

$$SSE = (w\varepsilon)^T \cdot (w\varepsilon) = \varepsilon^T w^2 \varepsilon = [Z - f(T) \cdot a]^T w^2 [Z - f(T) \cdot a]$$

Com isso, é possível encontrar o estimador de MQP de  $a$ , para o instante  $T$ ,  $\hat{a}(T)$ , bastando para tal diferenciar a função acima em relação ao vetor  $a$  e igualar a zero. Assim, chega-se a:

$$\hat{a}(T) = G^{-1}(T)g(T)$$

Sendo:  $G(T) = [w \cdot f(T)]^T [w \cdot f(T)]$ ;

$$g(T) = f^T(T)w^2Z.$$

Agora, fazendo a matriz de pesos  $w$  ser dada por:

$$w^2 = \begin{bmatrix} \beta^{T-1} & & & & 0 \\ & \beta^{T-2} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \beta \\ 0 & & & & & 1 \end{bmatrix}$$

Sendo  $0 < \beta < 1$ , que é considerado o fator de desconto.

Além disso, as funções  $f_i$  devem ser tais que a seguinte relação seja satisfeita:

$$f_i(t+1) = \sum_{j=1}^k L_{ij} f_j(t)$$

Sendo:  $i = 1, 2, \dots, k$ .

Com isso, o valor da função no instante  $t+1$  pode ser escrito como uma combinação linear das  $k$  funções no instante  $t$ , e as expressões dos estimadores e suas variâncias tornam-se simplificadas. Mais detalhes sobre essas equações podem ser encontrados em Montgomery (1990) e Lourenço (1998).

Existem várias aplicações nas quais foi utilizado o Método de Amortecimento Direto em previsões de séries de energia elétrica, usando dados de alta frequência.

Um exemplo do uso dessa técnica está em Quadrelli (1998), que a utilizou na previsão de carga elétrica horária e diária, com dados de uma concessionária do Brasil. Quadrelli ainda utilizou o método de decomposição de Gupta (que será apresentado ainda neste capítulo) para a previsão diária, e constatou que esse método apresentava melhores resultados do que o método de amortecimento direto. Já para a previsão dos dados horários, o método de amortecimento direto apresentou bons resultados. Mesmo tendo encontrado modelos que gerassem bons

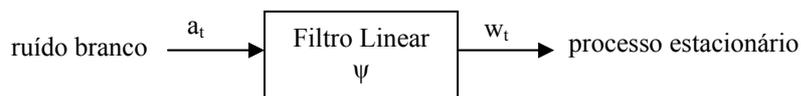
resultados para os períodos previstos, Quadrelli destacou que nenhum dos dois modelos mostrou um bom ajuste ao prever um dia de feriado.

Um outro exemplo do uso dessa metodologia pode ser encontrado em Esteves (2003), em que a autora utilizou este modelo e o de Holt-Winters com múltiplos ciclos para fazer previsão horária, 24 passos à frente, da carga elétrica de uma concessionária do Brasil. No entanto, o modelo de amortecimento direto não apresentou bons resultados ao ser comparado ao outro modelo utilizado.

## 2.6 – Modelo de Box & Jenkins

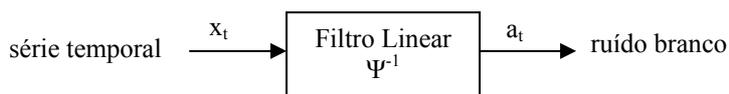
O modelo de Box & Jenkins (B&J) surgiu na década de 1970 e tem como base a Teoria Geral de Sistemas Lineares, a qual supõe que a passagem de um ruído branco por um filtro linear de memória infinita gera um processo estacionário de segunda ordem (Souza e Camargo, 2004). Assim, tem-se que, como na figura 2.02:

Figura 2.02 - Geração de uma série temporal



Então, o objetivo é encontrar o sistema inverso que, a partir de uma série temporal, seja capaz de gerar um ruído branco, como a figura 2.03 abaixo. Se isto ocorrer, é porque se conseguiu captar toda a estrutura de dependência da série temporal.

Figura 2.03 - Análise de uma série temporal



Essa metodologia assume que a série temporal de interesse foi gerada por um processo estacionário de segunda ordem. Como nem sempre isto ocorre, aplicam-se diferenciações sucessivas até torná-la estacionária. Assim, na

formulação B&J, utiliza-se o operador de atraso “B” (*backward shift operator*) definido como:

$$B^k Z_t = Z_{t-k}$$

O modelo de B&J segue a equação:

$$w_t = \Psi(B)a_t = \sum_{k=0}^{\infty} \psi_k a_{t-k}$$

Como  $\Psi(B)$  possui infinitos parâmetros, o que é um problema, B&J argumenta que, sob algumas restrições, pode-se dizer que todo polinômio infinito pode ser expresso pelo quociente de dois polinômios finitos. Sendo assim:

$$\Psi(B) = \frac{\theta(B)}{\phi(B)}$$

Sendo:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad \Leftrightarrow \text{Polinômio MA}(q)$$

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad \Leftrightarrow \text{Polinômio AR}(p)$$

Assim, tem-se os modelos ARMA(p,q) que possuem a seguinte forma:

$$\phi(B)w_t = \theta(B)a_t$$

Como a série temporal precisar ser estacionária para poder ser modelada, utiliza-se o operador diferença a fim de diferenciar a série quantas vezes forem necessárias para torná-la estacionária de segunda ordem. Dessa forma, dada uma série  $Z_t$  não estacionária, quer-se torná-la em uma série  $X_t$  estacionária, então:

$$X_t = Z_t - Z_{t-1} = Z_t - BZ_t = (1 - B)Z_t = \nabla Z_t$$

Sendo:  $\nabla = (1 - B) \Leftrightarrow$  operador de diferença

Generalizando, uma série  $w_t$  torna-se estacionária após a aplicação de “d” ( $d=0,1,2,\dots$ ) diferenças na série original  $Z_t$ :  $w_t = \nabla^d Z_t$ . Com isso, têm-se os modelos ARIMA(p,d,q) da seguinte forma:

$$\phi(B)\nabla^d Z_t = \theta(B)a_t$$

### 2.6.1- Modelos Box & Jenkins com Sazonalidade: SARIMA

É comum encontrar séries, além de não estacionárias, que apresentam a componente sazonal. Por isso, B&J desenvolveram os modelos SARIMA, os

quais incluem a correlação serial dentro e entre os períodos sazonais. Neste caso, a equação do modelo segue:

$$\phi(B)\Phi(B^S)\nabla_S^D\nabla^d Z_T = \theta(B)\Theta(B^S)a_t$$

Sendo:

$\phi(B)$ : operador não sazonal auto-regressivo;

$\phi_i$ : parâmetros auto-regressivo não-sazonais;

$\nabla^d = (1 - B)^d$ : operador diferença não sazonal de ordem d;

$\Phi(B^S)$ : operador sazonal auto-regressivo;

$\Phi_i$ : parâmetros auto-regressivo sazonais;

$\nabla_S^D = (1 - B^S)^D$ : operador diferença sazonal de ordem D;

$\theta(B)$ : operador não sazonal de médias móveis;

$\theta_i$ : parâmetros de médias móveis não sazonais;

$\Theta(B^S)$ : operador sazonal de médias móveis

$\Theta_i$ : parâmetros de médias móveis sazonais.

O modelo com essa estrutura é denominado SARIMA(p,d,q)x(P,D,Q)<sub>S</sub> e o procedimento para a obtenção desse modelo segue os mesmos passos que a modelagem ARIMA, usada para séries não sazonais.

Pode-se ainda escrever o modelo SARIMA em função de dois ciclos, isto é SARIMA(p,d,q)x(P<sub>1</sub>,D<sub>1</sub>,Q<sub>1</sub>)<sub>S1</sub>x(P<sub>2</sub>,D<sub>2</sub>,Q<sub>2</sub>)<sub>S2</sub>, de forma que possa ser usado em dados de alta frequência. Assim tem-se:

$$\phi_p(B)\Phi_{P_1}(B^{S_1})\Omega_{P_2}(B^{S_2})\nabla^d\nabla_{S_1}^{D_1}\nabla_{S_2}^{D_2}Z_T = \theta_q(B)\Theta_{Q_1}(B^{S_1})\Psi_{Q_2}(B^{S_2})a_t$$

Um exemplo da aplicação desta metodologia pode ser encontrado em Taylor & McSharry (2007), no qual este modelo (além de outros) é aplicado para séries de energia elétrica de onze países da Europa, para a previsão horária, vinte e quatro passos à frente.

## 2.7 – Modelo de Decomposição de Gupta

Este modelo foi desenvolvido para ser utilizado em previsão de curto prazo e para tal utiliza os dados históricos e os efeitos das variáveis climáticas, como será apresentado a seguir (Bunn & Farmer, 1985).

### 2.7.1 – Modelo para a Carga Padrão

Para formular a carga padrão, utiliza-se um método estocástico de forma que a carga real, da hora “h” do dia “d”, é modelada através de um componente padrão e um residual. Sua forma aditiva pode ser representada por:

$$Z(d, h) = P(d, h) + X(d, h)$$

Sendo:  $Z$  = a carga da hora  $h$ , para o dia  $d$ ;

$P$  = componente padrão;

$X$  = componente residual que contém variações na hora devido a fatores aleatórios.

A componente padrão pode ser dividida em duas partes, uma que representa o nível e a outra o efeito do ciclo semanal, assim:

$$P(d, h) = N(d, h) + S(d, h)$$

Sendo:  $N$  = componente do nível da hora  $h$  para o dia  $d$ ;

$S$  = componente do ciclo semanal (representa o efeito do dia da semana) para a hora  $h$ , do dia  $d$ .

Então:

$$Z(d, h) = N(d, h) + S(d, h) + X(d, h)$$

Neste caso, a componente do nível é praticamente uma constante, variando muito pouco de um dia para o outro; a componente do ciclo reflete o padrão semanal da carga; a componente residual, a variação na carga horária devido a fatores aleatórios.

A seguir é mostrado como cada uma das componentes acima é calculada:

- A componente do nível é calculada por uma média móvel de tamanho 7:

$$N(d, h) = \frac{1}{7} \sum_{i=d-6}^d Z(i, h), \text{ para } d \geq 7$$

- A componente do ciclo semanal é calculada por um amortecimento exponencial:

para  $i \geq 7$

para  $i < 7$

$$S(d, h) = \begin{cases} S(d-7, h) + \alpha[Z(d, h) - N(d, h) - S(d-7, h)] \\ Z(d, h) - N(7, h) \end{cases}$$

Sendo  $\alpha \in [0, 2; 0, 5]$

- A componente aleatória pode ser modelada por um modelo auto-regressivo:

$$X(d, h) = AX(d-1, h) + E(d-1, h)$$

Sendo: A uma matriz 24x24 que possui os coeficientes;

E um vetor 24x1 que possui os erros do modelo

Assim, o modelo de previsão é dado por:

$$\hat{Z}(d+1, h) = N(d, h) + S(d, h) + \hat{X}(d+1, h)$$

Sendo:  $\hat{X}(d+1) = AX(d)$ , estando aqui X representado na forma matricial.

### 2.7.2 – Modelo para as variáveis climáticas

Para modelar as variáveis climáticas, Gupta propôs um modelo no qual a essência é parecida com a do modelo para a carga padrão. Neste modelo, o pico de carga do dia será representado em termos das variáveis climáticas:

$$W(d) = B(d) + C(d) + M(d) + A(d)$$

Sendo:  $W(d)$  = o pico da carga do dia d;

$B(d)$  = a componente de nível do pico da carga do dia d;

$C(d)$  = a componente do comportamento semanal do pico da carga do dia d;

$M(d)$  = a componente meteorológica do pico da carga do dia d;

$A(d)$  = a componente aleatória do pico da carga do dia d.

No modelo acima, a componente meteorológica pode representar uma função linear ou uma transformação de variáveis climáticas.

A equação de previsão desse modelo é:

$$\hat{W}(d+1) = B(d+1) + C(d+1) + M(d+1)$$

Com isso, a equação de previsão da carga final é dada pela combinação desses dois modelos:

$$P(d+1, h) = F[\hat{Z}(d+1, h), \hat{W}(d+1)]$$

Um uso dessa metodologia está em Sobral (1999), em que a autora desenvolve um modelo de previsão de curto prazo para carga elétrica horária, empregando, também, informações climáticas. Nesse trabalho, as curvas típicas

de carga foram classificadas usando o mapa auto-organizável de Kohonen. Assim, o modelo de previsão de carga de Gupta foi adaptado a essas entradas. Uma outra alteração que a autora propôs a este modelo foi o uso de um sistema de lógica fuzzy para tratar as variáveis meteorológicas ao invés de utilizar o modelo de carga-clima proposto por Gupta.

Este modelo também foi utilizado por Quadrelli (1998), como descrito na seção de Amortecimento Direto, e apresentou bons resultados.

## 2.8 – Regressão Dinâmica

Os modelos de regressão linear geralmente utilizados partem do princípio que os erros oriundos do modelo possuem média zero, variância constante, distribuição Normal e independência.

Contudo, sabe-se que os resíduos gerados pelos modelos de séries temporais costumam apresentar correlações positivas, e mais: erros positivos têm a tendência de serem seguidos por outros também positivos, o mesmo ocorrendo no caso dos erros negativos.

De acordo com Barros e Souza (1995), ao modelar uma série temporal utilizando um modelo de regressão, os ruídos não podem ser considerados independentes e isso gera conseqüências nos testes e nos resultados gerados pelos modelos de regressão. Uma das conseqüências, por exemplo, da autocorrelação dos resíduos é que os estimadores de mínimos quadrados, apesar de serem não tendenciosos, não têm variância mínima. Além disso, os estimadores da variância e dos erros padrões dos coeficientes são subestimados (dando a impressão de que são mais precisos do que na realidade), o que invalida os intervalos de confiança e os testes de hipótese.

Por isso, para séries temporais, os modelos de regressão dinâmica são mais utilizados do que os modelos de regressão linear, uma vez que estendem os modelos usuais de regressão ao levarem em conta a hipótese de que os erros não são independentes.

Nos modelos de regressão dinâmica, a dinâmica das séries temporais e o efeito das variáveis explicativas são combinados. Por isso, esses modelos devem ser usados quando existe uma estrutura de dependência entre a variável

dependente e as variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série a ser explicada mostrar que não é possível supor a independência dos erros.

Os modelos de regressão dinâmica podem ser descritos pela equação:

$$\varphi(B)Y_t = \beta X_t + \varepsilon_t$$

Sendo:

$Y_t$  = variável dependente (endógena) no instante  $t$ ;

$\beta$  = vetor de coeficientes das variáveis causais, estimado por mínimos quadrados;

$X_t$  = vetor de variáveis causais (exógenas)<sup>1</sup> no instante  $t$ ;

$\varepsilon_t$  = ruído aleatório associado ao modelo;

$\varphi(B)$  = polinômio autoregressivo de ordem  $p$ , isto é:  $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$  sendo  $B$  o operador de atraso.

Nessa estrutura de regressão dinâmica, pode-se considerar  $X_t$  como sendo as variáveis causais em  $t$  e também em períodos defasados. Já o polinômio  $\varphi(B)$  é que permite a existência de defasagens da variável dependente, isto é, se  $\varphi(B)=1$ , não existe defasagem na variável dependente (neste caso, apenas as variáveis exógenas influenciam o modelo), mas se  $\varphi(B) \neq 1$ , defasagens na variável dependente fazem parte do modelo.

A estratégia mais comum que é utilizada na construção de um modelo de regressão dinâmica é a *bottom-up*, na qual inicia-se com um modelo simples e vai refinando-o, incluindo novas variáveis, até encontrar o modelo adequado.

Vale lembrar que em um modelo de regressão dinâmica, as previsões dependem não apenas do histórico da série, mas também dos valores previstos para as variáveis causais. Isto caracteriza um aspecto importante que é a elaboração de cenários para as variáveis causais, gerando flexibilidade na análise das previsões.

Para dados de alta frequência pode-se utilizar a regressão dinâmica fazendo um modelo para cada periodicidade. Por exemplo, para prever dados de quinze minutos, ter-se-ia que estimar 96 modelos, um para cada quinze minutos.

---

<sup>1</sup> As variáveis exógenas representadas por  $x_t$  têm seus valores determinados “fora” do modelo de regressão, enquanto a variável endógena  $Y_t$  é determinada a partir das variáveis exógenas e do ruído  $\varepsilon_t$ .

Este tipo de modelagem foi usado por Cancelo, Espasa & Grafe (2007), ao tratar dados de energia elétrica da Espanha. No trabalho em questão, o objetivo era fazer previsão para dados horários e, portanto, foram utilizadas vinte e quatro equações, sendo um modelo para cada hora. Nesse trabalho, os autores utilizaram além do passado da própria série, informações sobre temperatura.

Um outro trabalho que também utilizou a regressão dinâmica foi o de Ramanathan *et all* (1997), que tinha o objetivo de modelar dados horários de energia elétrica e, para tanto, foi utilizado um modelo de regressão para cada hora, sendo um modelo diferente para os finais de semana, totalizando assim quarenta e oito modelos. Nesse também foram utilizadas informações sobre temperatura.

## 2.9 - Redes Neurais

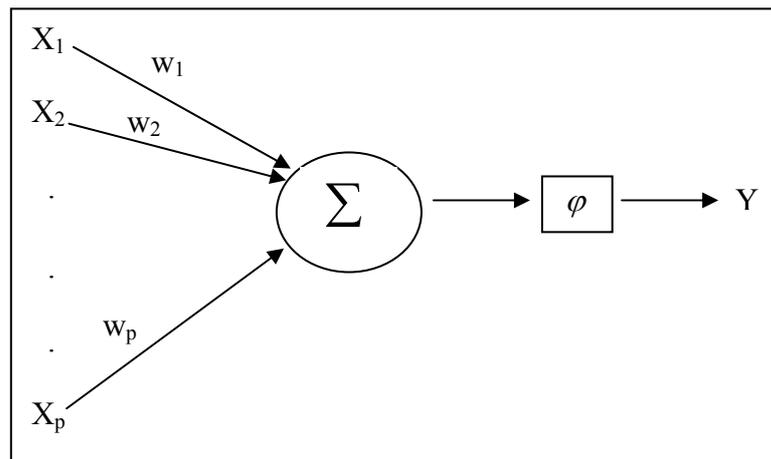
A rede neural (RN) é um sistema computacional que surgiu tomando como base o funcionamento do cérebro humano. Sua motivação inicial era a de realizar tarefas que o cérebro executa, como por exemplo, o reconhecimento de padrões, percepção e controle motor (Carvalho *et all*, 1998).

Uma vantagem do uso da RN é a sua capacidade de resolver problemas sem precisar da definição de regras ou de modelos explícitos. Assim, é possível trabalhar com situações onde não é possível criar um modelo da realidade ou situações que possuem mudanças frequentes no ambiente (Zanini, 2000).

A RN é formada por dois elementos básicos: um neurônio artificial, que é o processamento, e o peso ou sinapse, que é a conexão. Assim a RN será formada por um conjunto de neurônios artificiais que são conectados entre si por um conjunto de pesos. A figura 2.04 representa o modelo de um neurônio artificial (Haykin, 1999).

Dessa forma, pode-se sintetizar o funcionamento do neurônio artificial da seguinte forma: as entradas são multiplicadas por seus respectivos pesos, em seguida esses resultados são somados e depois submetidos à função de ativação, resultando na saída.

Figura 2.04 - Modelo de um neurônio artificial



A modelagem em RN é esquematizada da seguinte forma: escolha das observações para serem usadas no treinamento e das que serão usadas na validação, escolha da arquitetura apropriada e escolha do algoritmo de treinamento da rede.

Depois de especificada essa metodologia de modelagem, a rede está apta a “adquirir” o conhecimento. Esse conhecimento ocorre através do processo de treinamento (aprendizado), no qual a informação é armazenada como pesos, sendo que a combinação deles, de forma não linear, produz o conhecimento.

O tipo do aprendizado é determinado pela forma como a adaptação dos pesos (parâmetros) é feita, o que pode ocorrer de duas formas:

- a) Aprendizado Supervisionado: são fornecidos à rede pares de entradas e saídas já ocorridas, de forma que a rede possa encontrar a ligação entre os pares de entrada e saída fornecidos.
- b) Aprendizado Não Supervisionado: são fornecidos apenas os padrões de entrada, com os quais a rede estabelece características semelhantes que a torna capaz de criar classes ou grupos automaticamente.

Uma RN trabalha em duas fases distintas (tendo um conjunto de observações distinto para cada uma das fases): o treinamento e os testes. Na primeira parte, a rede tem que adquirir o conhecimento. Na segunda, é avaliado seu poder de generalização, que será importante seja para fazer, classificação ou projeção.

Uma característica importante da RN é que quanto mais simples a arquitetura da rede (número de camadas escondidas e número de neurônios em

cada camada), maior será seu poder de generalização, pois uma rede excessivamente complexa irá captar todos os detalhes do conjunto de treinamento, ao invés de aprender apenas as propriedades gerais.

A RN, em alguns casos, tem desempenho superior aos modelos estatísticos quando aplicada a problemas não lineares. Em casos lineares, geralmente, os métodos estatísticos tradicionais costumam ser mais eficientes.

Para a previsão de curto prazo para carga elétrica, tem-se uma aplicação em Darbellay & Slama (2000), além disso, podem-se encontrar mais detalhes em Hippert, Pedreira e Souza (2001), onde há uma revisão sobre vários artigos que utilizam a RN.

Pode-se ainda usar a RN modelando uma rede para cada dia a ser previsto ou ainda uma rede para alguns dias, como feito por Lazo *et al* (2006), onde era necessário prever quinze minutos para 15 dias à frente para dados de demanda de energia elétrica e para tal foram utilizadas oito redes neurais. Assim, foi utilizada uma rede para prever os períodos do primeiro dia à frente, uma segunda rede para prever o segundo e o nono dia à frente, uma terceira para fazer a previsão de todos os períodos do terceiro até o décimo dia à frente e assim por diante, até chegar à oitava rede que é responsável por fazer a previsão dos períodos do oitavo e do décimo quinto dia à frente.

Também pode ser feito como em McMenemy & Monforte (1998), que tinham o objetivo de prever os dados horários do dia e, para isto, utilizaram vinte e quatro redes neurais, isto é, uma para cada hora.

Há ainda a alternativa de aplicar um modelo híbrido. Este tipo de modelo foi usado por Rizzo (2001), cujo modelo era composto por um modelo de amortecimento exponencial (no qual ele tinha a idéia de captar as características lineares da série) e por um modelo de redes neurais do tipo *feed-forward* (com o qual ele visava captar as características não lineares da série). Nesse trabalho, Rizzo tinha o objetivo de prever dados de energia elétrica a cada trinta minutos para quinze dias à frente. Para tanto, ele fazia a previsão para os períodos de 't' até 't+6' (sendo 't' o primeiro dia a ser previsto). Em seguida essa previsão era utilizada como entrada para prever os dias 't+7' até 't+13', sendo que esses eram também usados como entradas para prever 't+14' e 't+15'.

## 2.10- Lógica Fuzzy

A Lógica Fuzzy objetiva imitar a habilidade dos seres humanos de tomar decisões em um ambiente de incerteza e imprecisão. Isso ocorre através da manipulação, simultânea, de dados numéricos e do conhecimento lingüístico. Assim a lógica fuzzy realiza um mapeamento não linear de valores numéricos de entrada em valores numéricos de saída, sendo esse mapeamento estabelecido através da teoria de conjuntos nebulosos e da lógica nebulosa (Mendel, 1995). Na figura 2.05 há um esquema do sistema de lógica fuzzy.

Este sistema é composto por quatro módulos:

- Fuzzificador: tem a função de ativar as regras que estão associadas às variáveis lingüísticas, isto é, mapeia valores numéricos em conjuntos nebulosos.
- Banco de Regras: é fornecido por especialistas ou extraídas de dados numéricos, sendo formado por regras do tipo SE – ENTÃO.
- Inferência: determina como as regras são combinadas a fim de gerar a saída fuzzy. Assim, este módulo tem a função de mapear conjuntos fuzzy de entrada em conjuntos fuzzy de saída.
- Defuzzificador: mapeia o conjunto fuzzy de saída em um valor numérico.

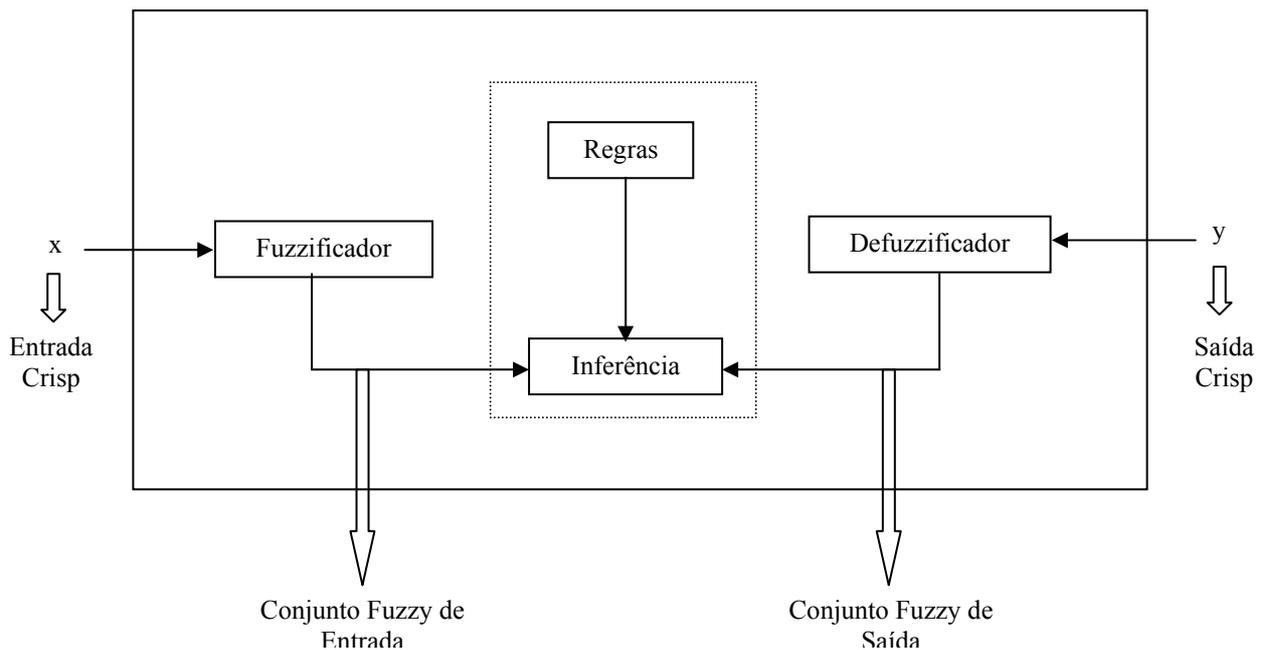
As regras do sistema fuzzy, definidos da forma SE-ENTÃO, envolvem variáveis lingüísticas que são agregadas utilizando conectores lógicos do tipo E/OU. Uma regra pode ser exemplificada como:

SE  $x_1$  é quente E  $x_2$  é muito baixo ENTÃO y gira um pouco para a direita

Nessa regra,  $x_1$ ,  $x_2$  e y são variáveis lingüísticas, possuindo valores dentro do conjunto fuzzy, do tipo quente, muito baixo e um pouco para a direita. Cada regra que foi ativada irá resultar em um conjunto fuzzy que varia de acordo com o grau de disparo da regra.

No sistema fuzzy para previsão de séries, as regras são extraídas a partir dos dados numéricos, o que pode ocorrer por duas formas distintas: os dados da série estabelecem os conjuntos fuzzy que irão formar os antecedentes e conseqüentes das regras, ou, primeiro, são pré-estabelecidos os conjuntos fuzzy para os antecedentes e os conseqüentes e, depois, os dados são associados a estes conjuntos.

Figura 2.05 - Sistema de Lógica Fuzzy



Um exemplo de aplicação da lógica fuzzy em previsão pode ser encontrado em Lourenço (1998) e em Mori & Kobayashi (1996), que utilizaram a lógica fuzzy para a previsão de carga elétrica de curto-prazo. Mais detalhes sobre essa metodologia, podem ser encontrados em Mendel (1995).

Pode-se ainda utilizar sistemas neuro-fuzzy, que consistem em uma arquitetura híbrida, na qual as formas de aprendizado da rede neural são combinadas com um sistema de inferência fuzzy. Dessa forma, é possível combinar a característica de aprendizado das redes neurais com a interpretabilidade dos sistemas fuzzy. Uma das formas de utilizar esse sistema neuro-fuzzy é através do ANFIS, que está detalhado em Jang (1993), e uma aplicação desse método para previsão de curto prazo de energia elétrica pode ser encontrado em Papadakis *et al* (1998) e Bakirtzis *et al* (1995).