

4. Modelagem Proposta

Uma das conseqüências da privatização dos mercados de energia foi a maior ênfase que passou a ser dada à previsão da curva de carga, isto é, o valor da série de cargas em cada uma das 24 horas de um dia futuro (ou em cada uma das 48 meias-horas), já que estas previsões servem de base para a fixação do preço de energia a cada momento do dia. Diversas dificuldades são trazidas ao pesquisador, criadas principalmente pela autocorrelação das séries (como prever o valor da variável na hora h do dia seguinte, se este depende do valor na hora $(h-1)$ do mesmo dia, também desconhecido?) e pela alta dimensionalidade dos problemas (pois as curvas de carga podem ser encaradas como vetores de 24 dimensões).

A resolução para a previsão das 24 horas de carga, nesta tese, será dada pela combinação entre a previsão do perfil do dia em PU (por unidade) com a previsão da demanda média diária.

A idéia inicial de combinar os dois resultados surgiu durante o projeto de P&D (Pesquisa e Desenvolvimento), realizado entre a PUC-Rio (Pontifícia Universidade Católica do Rio de Janeiro) e a LIGHT Serviços de Eletricidade S.A (Souza et al., 2001).

Com isto, o método de previsão de carga horária basea-se na combinação de duas componentes: a previsão do perfil em PU e a previsão de demanda média. A previsão final é dada pela equação:

$$\hat{C}_{i,h} = \hat{D}_i \times \hat{P}_{i,h} \quad (4)$$

onde:

$\hat{C}_{i,h}$: carga final prevista no dia i e hora h

\hat{D}_i : demanda média prevista no dia i

$\hat{P}_{i,h}$: perfil previsto em PU no dia i e hora h

O modelo proposto adota, na obtenção de suas componentes, as seguintes etapas:

Tabela 1 - Descrição das Etapas da Modelagem Proposta

1ª Etapa	<p style="text-align: center;"><u>Classificação do Perfil em PU</u></p> <p><i>Metodologia</i> : SOM (Mapa Auto Organizável de Kohonen)</p> <p><i>Entrada</i> : Perfil em PU das 24 horas</p> <p><i>Saída</i> : Não possui (Procedimento Não-Supervisionado)</p>
2ª Etapa	<p style="text-align: center;"><u>Discriminação do Perfil em PU</u></p> <p><i>Metodologia</i> : Árvore de Decisão</p> <p><i>Entrada</i> : Temperatura mínima e máxima do dia anterior; Razão da média das cargas das cinco últimas horas com a média das cargas das cinco primeiras horas do dia anterior</p> <p><i>Saída</i> : Grupo dos perfis típicos</p>
3ª Etapa	<p style="text-align: center;"><u>Previsão da Demanda Média</u></p> <p><i>Metodologia</i> : Árvore de Regressão</p> <p><i>Entrada</i> : Perfil em PU do dia (24 horas)</p> <p><i>Saída</i> : Demanda Média do dia</p>

A figura 3 apresenta o esquema da modelagem proposta:

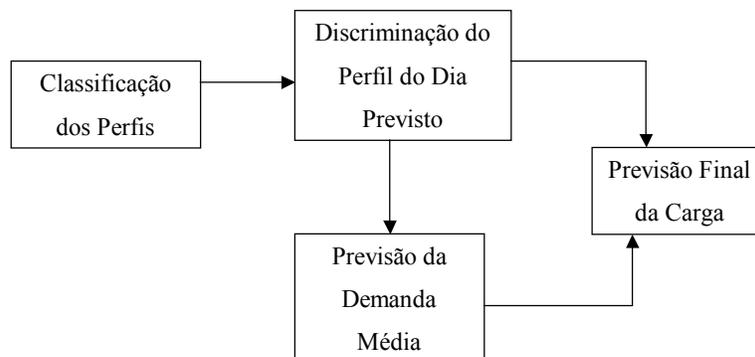


Figura 3 - Esquema da Modelagem Proposta

Nas próximas seções serão detalhados os procedimentos empregados em cada etapa da modelagem proposta.

4.1.

Etapa I: Classificação do Perfil em PU

Na etapa I da modelagem proposta é realizada a classificação dos perfis diários típicos. A classificação destes perfis possibilita o rompimento da barreira da classificação rígida por estação do ano. Além disso, a realização da classificação permite que o modelo proposto respeite as peculiaridades dos perfis característicos de cada concessionária (por exemplo, uma concessionária do nordeste do Brasil, provavelmente, apresenta perfis diferentes de uma concessionária do sul do país).

Nessa tese, o Mapa Auto-Organizável de Kohonen (SOM) foi empregado como procedimento de classificação. O SOM é um tipo de rede de neural e está descrito no Anexo III.

Deve-se mencionar que o algoritmo do Mapa Auto-Organizável de Kohonen não foi originalmente proposto para problemas de classificação. Embora se encontre na literatura diversas aplicações neste campo (Hsu e Yang, 1991), este algoritmo não produz um protótipo centralizado nos agrupamentos, nem tão pouco otimiza a compactação destes. Alguns esquemas híbridos que aplicam técnicas estatísticas clássicas em conjunto com o algoritmo de Kohonen produziram uma melhora considerável na qualidade de classificação em outras áreas (Medeiros, 1999). Uma outra alternativa para classificação é o emprego de técnicas estatísticas clássicas combinadas com redes hierárquicas aglomerativas (Velasquez, et al., 2001).

4.2.

Etapa II: Discriminação do Perfil em PU

Após a classificação é realizada a segunda etapa da modelagem proposta: a discriminação do perfil em PU do dia previsto. Ou seja, uma vez estabelecido um conjunto de perfis típicos para uma determinada concessionária, serão empregados atributos meteorológicos (temperatura) e não-meteorológicos (carga) para prever o perfil em PU do dia a ser previsto.

Em Sobral (1999) foi empregado como atributos não-meteorológicos, a carga média das 5 primeiras horas e a carga média das 5 últimas horas do dia a ser

previsto e como atributos meteorológicos, a temperatura média das 5 primeiras horas e a temperatura média das 5 últimas horas do dia a ser previsto. No entanto, a solução meteorológica empregada não é totalmente realística do ponto de vista brasileiro, já que as concessionárias brasileiras não apresentam as previsões de temperaturas horárias para o dia seguinte.

Convém destacar aqui, as dificuldades gerais em empregar previsão de temperatura como entrada do modelo de previsão:

- A qualidade das previsões meteorológicas, em geral, deixa a desejar;
- Previsões meteorológicas podem não ser disponíveis, ou serem muito caras;
- Os modelos de previsão de carga encontrados na literatura, na grande maioria das vezes, não avaliam os “erros totais” que levam em consideração os erros originários da previsão de temperatura e os erros originários do modelo de previsão proposto. São empregados valores reais de temperatura no lugar das temperaturas previstas para avaliar o desempenho dos modelos propostos.

Com isto, uma solução meteorológica mais adequada à realidade brasileira foi adotada nesta tese. A solução foi o emprego de valores diários passados de temperaturas máximas e mínimas (e não temperaturas horárias previstas).

Em relação aos atributos não-meteorológicos, foi empregada a razão da carga média das 5 primeiras horas (MADRUGADA) e a carga média das 5 últimas horas (NOITE).

A idéia de utilizar a carga média da madrugada e da noite é fornecer o sentido de variação da carga (crescimento ou diminuição da carga). Uma situação onde tem-se somente temperaturas máximas e mínimas não é suficiente para dar, sempre, uma boa previsão de carga, já que a temperatura máxima ocorre entre 13:00 e 14:00 horas e a mínima pode ocorrer no fim da madrugada (na maioria dos casos) ou a noite (se a temperatura do dia cair) ou no fim da tarde (caso de inversão térmica). Com isto, as cargas médias da madrugada e da noite complementam as informações de temperaturas mínimas e máximas diárias.

Por fim, a idéia da razão das cargas médias da madrugada e da noite é tornar a série estacionária, o que possibilita o emprego de dados passados na modelagem sem prejudicar o procedimento de discriminação/previsão de perfis em PU.

Nesta etapa, como procedimento para discriminação do perfil do dia previsto, empregou-se árvore de decisão, mais especificamente o CART (Breiman, 1983). Já que a árvore de decisão possui treinamento supervisionado,

foram estabelecidos como saída desta árvore os grupos de perfis típicos obtidos pelo Mapa Auto-Organizável de Kohonen (Classificação - Etapa I).

4.3.

Etapa III: Previsão da Demanda Média

Na etapa III da modelagem proposta é obtida a demanda média prevista através de árvore de regressão. As entradas e a saída desta árvore de regressão são estabelecidas como:

ENTRADA: perfil em PU do dia previsto (24 horas), já que temos a discriminação do grupo que pertencerá o dia previsto (obtido na Etapa II da modelagem proposta).

⇒ Vamos supor que a Etapa II da modelagem proposta identifica que amanhã será um dia do grupo 1. Então, emprega-se a média dos perfis em PU desse grupo e a utiliza como entrada da árvore de regressão. Lembrando que, o banco de dados de cada grupo é atualizado sempre que um dia desse grupo ocorre.

SAÍDA: demanda média do dia previsto.

A árvore de regressão utilizada é o TS-TARX (Tree Structured - Threshold Autoregression with eXternal Variables), implementado inicialmente para 5 variáveis de entrada por Aranha e Veiga (2001). Nesta tese, o procedimento TS-TARX foi implementado para mais de 5 variáveis de entrada, no nosso caso, 24 variáveis de entrada (perfil em PU das 24 horas).

O TS-TARX (“Tree Structured - Threshold Autoregression with eXternal Variables”) é uma extensão do modelo TAR. No entanto, neste trabalho, os nós folhas não apresentam modelos Autoregressivos e, sim, modelo de regressão múltipla.

Como modelo autoregressivo AR (p) tem-se:

$$W_t = a_t + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} \quad (5)$$

O nome Autoregressivo se deve ao fato de que W_t no instante t é função dos W 's nos instantes anteriores.

Já como modelo de regressão múltipla tem-se:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i \quad (6)$$

onde:

$\beta_0, \beta_1, \dots, \beta_{p-1}$ são parâmetros

$X_{i,1}, X_{i,2}, \dots, X_{i,p-1}$ são as variáveis conhecidas

ε_i são ruídos independentes $N(0, \sigma^2)$

$i=1, \dots, n$ (tamanho da amostra)

Y_i é a variável prevista

O TS-TARX apresenta duas fases: a primeira fase é um processo combinatório de regressão múltipla e a segunda fase é a essência do processo TS-TARX, onde a busca pelos limiares e quebras constituirá a árvore de regressão. Ao final da segunda fase, os resultados da primeira e da segunda fase são comparados e o melhor deles é escolhido.

Em linhas gerais, pode-se descrever o procedimento TS-TARX como:

- Na Fase I é realizada uma redução do modelo linear inicial de regressão múltipla, e o modelo vencedor dessa redução (com apenas variáveis significantes no modelo) será chamado de Modelo Reduzido. O BIC (Bayesian Information Criterion) calculado para o modelo reduzido será o critério de comparação com as combinações geradas pela árvore. Ou seja, sucessivas comparações de modelos através do critério BIC serão feitas, se o valor do BIC do modelo reduzido for menor entre todas as combinações geradas pela árvore, então o modelo será considerado linear e a árvore gerada não conterá ramos.
- Na Fase II é identificado o limiar para a quebra da árvore e é determinado o crescimento da mesma. O BIC dos dois modelos de regressão (nó da esquerda e nó da direita) deve ser menor que o BIC fornecido pelo modelo reduzido obtido na Fase I. Após a determinação deste limiar, criamos uma pergunta tal como: x_i é maior ou menor que este limiar? De modo que a sua resposta indicará um subconjunto de amostra, formando assim uma parte da árvore de decisão (crescimento).

Como a seguir será apresentado o procedimento TS-TARX, são convenientes as definições abaixo:

X - matriz de variáveis de entrada (ou independentes ou explicativas) de tamanho n por $(r+1)$;

n - tamanho da amostra;

r - número de variáveis de entrada;

Y - vetor da variável alvo de tamanho n ;

O modelo linear é definido algebricamente como $Y = X\beta + \varepsilon$ (descrição na equação 6).

Com intuito de ligar as definições de X e Y com o objetivo desta tese, temos no treinamento da árvore de regressão:

X é a matriz de variáveis de entrada que corresponde aos perfis horários

$$X = \begin{bmatrix} 1 & P_{1,1} & P_{1,2} & \cdots & P_{1,24} \\ 1 & P_{2,1} & P_{2,2} & \cdots & P_{2,24} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & P_{365,1} & P_{365,2} & \cdots & P_{365,24} \end{bmatrix}_{365 \times 25}$$

onde:

$P_{i,j}$: perfil em PU do dia i e hora j (variável de entrada)

n : tamanho da amostra (365 dias)

r : 24 (número de variáveis de entrada \Rightarrow PU das 24 horas)

Y é o vetor da variável alvo correspondente a demanda média

$$Y = \begin{bmatrix} \bar{D}_1 \\ \bar{D}_2 \\ \vdots \\ \bar{D}_{365} \end{bmatrix}_{365 \times 1}$$

onde:

\bar{D}_i : demanda média do dia i

Procedimento TS-TARX passo-a-passo:

1ª Fase: Determina modelo reduzido

1. Obtém-se a estimação por mínimos quadrados do modelo $Y=X\beta + \varepsilon$ (equação 6). Este modelo é chamado de Modelo Cheio (com todas as variáveis de entrada) durante o procedimento;
2. Faz-se todas as combinações possíveis de variáveis de entrada (modelo Reduzido) e obtém-se para cada combinação a estimação de mínimos quadrados e a estatística F. Na verdade, é realizado o Teste Simultâneo para os Coeficientes de Regressão (Neter, 1974).

$$F^* = \frac{SQE(\text{Reduzido}) - SQE(\text{Cheio})}{(n - p_R) - (n - p_C)} \div \frac{SQE(\text{Cheio})}{(n - p_C)} \quad (7)$$

onde

SQE (Cheio) : soma dos quadrados do erro do modelo cheio;

SQE (Reduzido) : soma dos quadrados do erro do modelo reduzido;

n : tamanho da amostra;

p_R : número de parâmetros estimados do modelo reduzido;

p_C : número de parâmetros estimados do modelo cheio.

3. Escolhe-se o modelo Reduzido de maior valor de estatística F;
4. Aplica-se o teste F para verificar a hipótese de nenhuma variável explicativa ser significativa no modelo reduzido escolhido.

Se o modelo selecionado for rejeitado pelo teste F, o modelo vencedor será a média das observações de Y.
5. Calcula-se o valor do critério de informação BIC (Bayesian Information Criterion) do modelo Reduzido.

$$BIC = \log(EQM) + \left(\frac{m+1}{n} \right) \log(n) \quad (8)$$

onde

EQM : erro quadrático médio do modelo reduzido;

m : número de parâmetros estimados do modelo reduzido;

n : tamanho da amostra;

Este BIC será chamado de $BIC_{SemQuebra}$.

A partir de agora, sucessivas comparações de modelos através do critério BIC serão feitas. Se o valor do BIC calculado para o modelo Reduzido for menor entre todas as combinações geradas pela árvore, então, o modelo é linear e a árvore gerada não conterá ramos.

2ª Fase: Determina limiar e quebra da árvore

Queremos um limiar tal que o BIC dos dois modelos de regressão (o modelo para o nó esquerdo e para o nó direito) seja menor que o BIC fornecido pelo modelo Reduzido (1ª Fase). Após a determinação do limiar, cria-se a pergunta: x_i é maior ou menor que o limiar? A resposta forma uma parte da árvore (crescimento da árvore).

6. Ordena-se a matriz de dados X por uma variável de entrada

A operação de ordenação, neste procedimento, é necessária para correr todas as tentativas de valores de “ponto de quebra” para a variável de entrada selecionada (x_i) de maneira eficiente. Precisamos tentar todos os valores de “ponto de quebra” para avaliar todas as divisões candidatas. As divisões candidatas modificam o conjunto de dados que irão para os lados esquerdo e direito, alterando assim o valor do erro quadrático médio em cada situação. Destaca-se que, o procedimento de ordenação da variável de entrada é realizado para todas (uma a uma) as variáveis de entrada.

7. Aplica-se o método de Mínimos Quadrados Recursivos (descrito no Anexo IV) na matriz de dados ordenada pela variável de entrada (x_k).

O método de Mínimos Quadrados Recursivos é utilizado para calcular as sucessivas regressões a medida em que a matriz ordenada é varrida, adicionando-se uma nova observação da amostra por vez. Assim, em cada iteração do algoritmo de partição recursiva, todas as divisões possíveis de cada variável de entrada são avaliadas e a melhor variação do erro é escolhida.

A dupla varredura (descrição no Anexo II), que é a aplicação dos mínimos quadrados recursivos na matriz de dados ordenada tanto em ordem crescente quanto em ordem decrescente, é importante para decidir o conjunto de dados que seguirá para o lado esquerdo e o conjunto de dados que seguirá para o lado direito da árvore. Ao ordenar os dados em ordem crescente tem-se a informação sobre o EQM do regime da esquerda e sua evolução devido a alteração do ponto de quebra por mínimos quadrados recursivos. Para obter informação do regime da direita, faz ordenação decrescente da matriz de dados em função da variável a ser explorada e realiza o mesmo procedimento descrito para o regime da esquerda.

8. Armazena-se o valor do EQM em cada iteração processada no item 7 para cada tamanho de amostra sucessiva.
9. Compõe-se o EQM para cada opção de variável decisória e suas possíveis quebras. O EQM é composto da seguinte maneira.

$$EQM_{\text{ComQuebra}} = \frac{n_e EQM_e + n_d EQM_d}{(n_e + n_d)} \quad (9)$$

onde

EQM_e : erro quadrático médio do modelo do regime da esquerda

EQM_d : erro quadrático médio do modelo do regime da direita

n_e : tamanho da sub-amostra do regime da esquerda

n_d : tamanho da sub-amostra do regime da direita

10. Escolhe-se o menor $EQM_{ComQuebra}$

11. Calcula-se o critério de informação BIC referente ao modelo com quebra

$$BIC_{ComQuebra} = \log(EQM_{ComQuebra}) + \frac{(m_e + m_d)}{(n_e + n_d)} \log(n_e + n_d) \quad (10)$$

onde

$EQM_{ComQuebra}$: erro quadrático médio do modelo com quebra

m_d : número de parâmetros estimados no regime da direita

m_e : número de parâmetros estimados no regime da esquerda

n_d : tamanho da sub-amostra do regime da direita

n_e : tamanho da sub-amostra do regime da esquerda

12. Compara-se o $BIC_{SemQuebra}$ (Fase I) com o $BIC_{ComQuebra}$ (Fase II).

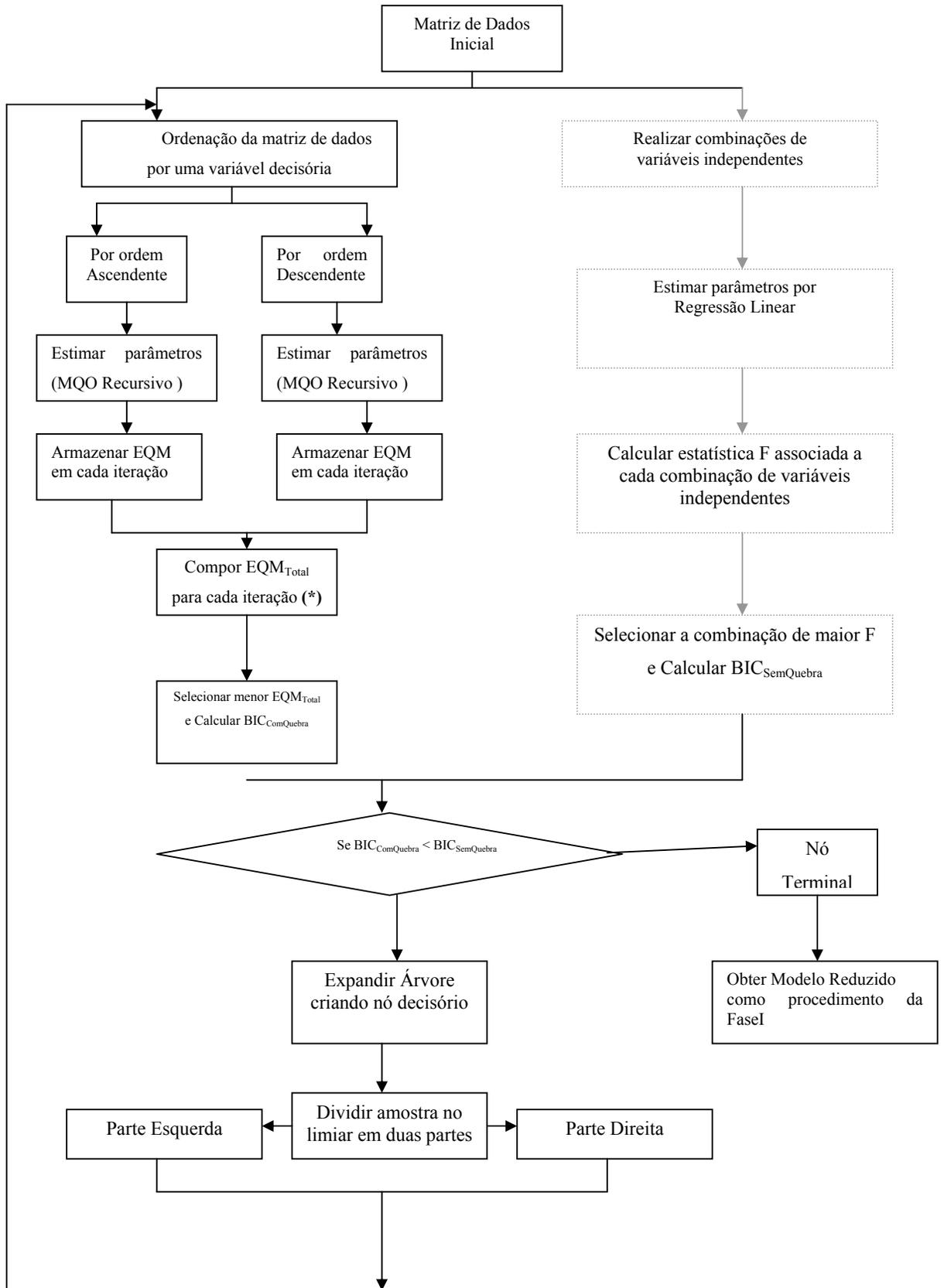
Se o $BIC_{ComQuebra} < BIC_{SemQuebra}$, a quebra é aceita e a árvore é expandida em dois novos nós (crescimento da árvore). A quebra da árvore representa não-linearidade significativa.

Se o $BIC_{ComQuebra} > BIC_{SemQuebra}$ é aceito o modelo linear (não há quebra) e o nó é rotulado como terminal (folha). Destaca-se que, sempre que um nó terminal é encontrado, aplica-se o processo de obtenção do modelo Reduzido como explicado na Fase I. Com isto, tem-se um modelo para cada regime (esquerda e direita):

$$Y_e = X_e \beta_e + \varepsilon_e$$

$$Y_d = X_d \beta_d + \varepsilon_d$$

A seguir tem-se o fluxograma do procedimento TS-TARX. Destaca-se que os retângulos pontilhados em cinza no fluxograma correspondem a Fase I do procedimento



(*) No momento de compor EQM_{Total} , a primeira iteração por ordem ascendente da variável decisória x_k corresponde a última iteração por ordem descendente dessa variável.

Figura 4 - Fluxograma da árvore de regressão TS-TARX