

5 Algoritmo para detecção de clientes potenciais

Conforme visto ao longo desta dissertação, o conhecimento do perfil do consumidor é fundamental para o incentivo ao uso de alternativas tecnológicas e avanço das pesquisas em eficiência energética, bem como para o planejamento energético das concessionárias e atendimento das metas de conservação de energia, necessárias e obrigatórias para o crescimento sustentável, de acordo com os programas e leis vigentes sobre este assunto no País.

Uma forma de orientar estas ações é através da análise da tendência das curvas de carga e dados históricos dos consumidores. A partir desta avaliação se torna possível atuar modulando estas curvas, através da aplicação do gerenciamento de demanda, no sentido de que se modifiquem e se ajustem visando à melhoria da eficiência.

Considerando que as concessionárias realizam medições em cada unidade consumidora para efeitos de faturamento da conta energia, logo é possível a partir desta massa de dados, construir as curvas de cargas de cada cliente e caracterizá-los de acordo com o consumo realizado diariamente, mensalmente e anualmente.

Conhecendo o perfil de consumo, se torna viável realizar estudos para incentivar a aplicação de sistemas de conservação de energia nos consumidores e conseqüentemente, otimizar o planejamento energético das concessionárias. Para tanto, precisa-se de ferramentas para a construção das curvas de carga e futura avaliação destas.

O algoritmo proposto neste trabalho trata os dados reais de medição da concessionária, mapeia as curvas de cargas de seus clientes e identifica os potenciais⁵ ao uso de tecnologias alternativas de eficiência energética.

Este recebe como entrada os dados reais de medição de consumo de energia de uma unidade consumidora e, a partir de critérios pré-estabelecidos em sua rotina, gera como saída a indicação dos clientes potenciais a sistemas de eficiência energética.

⁵ Clientes potenciais são os clientes que não possuem nenhum sistema alternativo de utilização de energia no horário de ponta (e.g.: gerador, termoacumulação, chillers).

O algoritmo foi elaborado visando ser um instrumento de apoio à decisão para as concessionárias, tanto em seu planejamento energético quanto na implantação de ações de eficiência energética.

Para a aplicação do algoritmo desenvolveu-se um sistema computacional gerador de curvas de carga, o SIMCAR e através deste, identifica os potenciais candidatos ao uso da termoacumulação, alternativa de eficiência energética foco desta dissertação, e simula as possibilidades de racionamento de energia para cada consumidor.

Ressalta-se, que sua aplicação é abrangente, permitindo a identificação e o fornecimento de subsídios de análise para quaisquer concessionárias e tecnologias de conservação de energia.

5.1. Construção do algoritmo

Para a construção do algoritmo de detecção, buscando fornecer uma ferramenta de apoio à decisão para as concessionárias, faz-se necessário adequar o algoritmo às características reais do problema.

No início da formalização do algoritmo são feitas as classificações e os agrupamentos, para uma filtragem dos clientes de interesse, num segundo momento são avaliadas as informações disponibilizadas nos levantamentos de dados realizados para este estudo, em seguida são realizados os tratamentos estatísticos dos dados.

5.1.1. Análise exploratória da população

Os clientes das concessionárias classificam-se em ramos e tarifas diferenciadas, conforme descrito no capítulo 2 deste trabalho, portanto faz-se necessária a adequada classificação e separação destes, para o estudo, o que é realizado através da análise exploratória da população.

A população selecionada foi a de grandes clientes, dos ramos de atividade comercial e industrial, notadamente do grupo A, com tarifas contratadas horosazonal verde ou azul, incluídos nos subgrupos AS, A3 e A4.

Nesta análise inicial filtrou-se 3.181, de cerca de 8.000 grandes clientes da concessionária, identificados como de interesse no estudo.

5.1.2. Levantamento da base de dados

O levantamento da base de dados é imprescindível, uma vez que é a principal informação de entrada do algoritmo de detecção dos clientes potenciais a conservação de energia.

O trabalho se desenvolveu com base nos dados reais de medições coletados, que para análise da sazonalidade da unidade consumidora foi de um ano.

Desta base foram extraídas as principais características das medições nela registrada, de forma a se obter elementos para a construção do algoritmo de detecção. Destaca-se, que alguns consumidores não apresentaram os registros dos doze meses almejados, nestes casos foram estudados os períodos disponíveis.

Nos relatórios com a memória de massa providos dos medidores apresentados no formato de Excel, constam as informações do equipamento de medição; as observações referentes a data da leitura; e o horário de demanda medida a cada quinze minutos.

Observou-se que em alguns horários não havia registros de medições, assim como algumas medições apresentavam valores discrepantes em relação as série de consumo típica, que podem ter ocorrido devido a diversos fatores, dentre eles: falha do equipamento de medição, queda de energia. Esta ocorrência é comum em massas de dados que armazenam um volume de informações muito grande.

Como estas inconsistências podem refletir no resultado final desejado, durante o mapeamento das curvas de cargas, averiguou-se o melhor método estatístico para o tratamento destes dados.

Todos os dados levantados foram importados para o sistema gerador de curvas de carga desenvolvido (SIMCAR), onde o algoritmo foi implementado, e a partir do qual foram realizados o tratamento dos dados, a geração das curvas de cargas reais de consumo e a detecção de potencialidade ao uso de termoacumulação dos clientes da concessionária.

5.1.3. Tratamento de dados aberrantes “*outlier*”

Para o tratamento dos dados aberrantes ou “*outlier*” foi utilizado o critério do intervalo de confiança do desvio padrão em torno da média dos dados,

comumente empregado para identificação de observações discrepantes, que se baseia em medidas pouco resistentes.

O intervalo de confiança (IC) é um intervalo estimado de um parâmetro de interesse de uma população, sendo usados para indicar a confiabilidade de uma estimativa. Em vez de estimar o parâmetro por um único valor, é dado um intervalo de estimativas prováveis. O quanto estas estimativas são prováveis será determinado pelo coeficiente de confiança $(1 - \alpha)$, para $\alpha \in (0,1)$. [24].

Uma medida de centralidade é um valor “típico” em torno do qual se situam os valores da variável e uma medida de dispersão para uma variável quantitativa é um indicador do grau de espalhamento dos valores da amostra em torno da medida de centralidade [44].

A equação 1 representa o intervalo de confiança utilizado no tratamento dos dados aberrantes.

$$(\bar{x} - 3. \sigma ; \bar{x} + 3. \sigma) \tag{1}$$

A aplicação deste critério aos dados de medição é realizada buscando no sistema SIMCAR, o histórico de medições do cliente, de um período de 30 dias, em que não tenha ocorrido dados faltantes, sendo então calculada as medias e as dispersões estimadas de todas as 96 medições de cada dia do mês, representada na matriz na equação 2.

$$\begin{array}{ccc} x_{1,1} & \cdots & x_{96,1} \\ \vdots & \cdots & \vdots \\ x_{1,30} & \cdots & x_{96,30} \\ \hline \bar{x}_1 & \cdots & \bar{x}_{96} \\ \hat{\sigma}_1 & \cdots & \hat{\sigma}_{96} \end{array} \tag{2}$$

Desta forma são encontrados os intervalos de confiança de cada uma das 96 medições, estabelecendo assim, os limites superiores e inferiores. Portanto, se uma medição for aberrante, esta ultrapassará as os limites do IC naquele horário medido.

O gráfico da figura 16 ilustra a curva (azul) das 96 medições reais, sem dados faltantes, realizadas em um dia útil de um hotel de grande porte, e as curvas (verde e vermelho), dos limites do intervalo de confiança destas

medições, sendo possível observar que nenhuma medição ultrapassou os limites do IC, logo neste caso não ocorreu dados aberrantes.

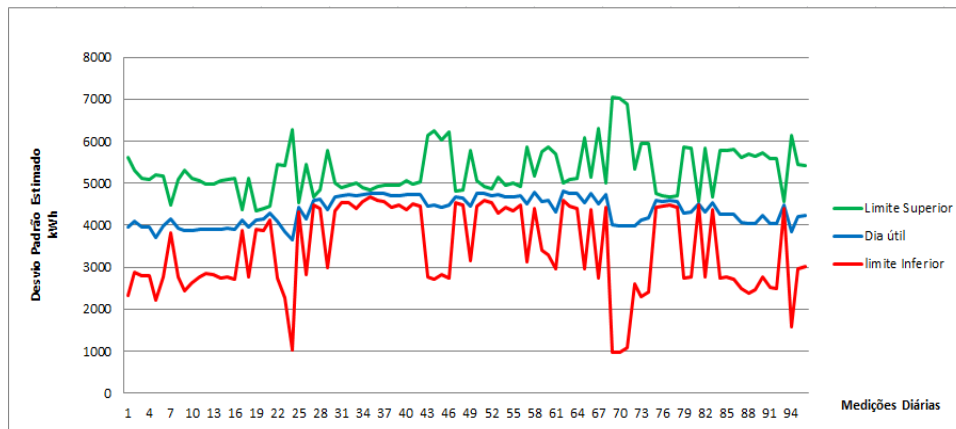


Figura 16 – Intervalo de confiança (96 medições de um dia útil)
 Fonte: o autor

Já a figura 17, ilustra a simulação da curva anterior, considerando a existência de dados aberrantes, permitindo assim verificar, que em algumas medições, a curva azul ultrapassa as verde e vermelha, caracterizando a existência de medições aberrantes.

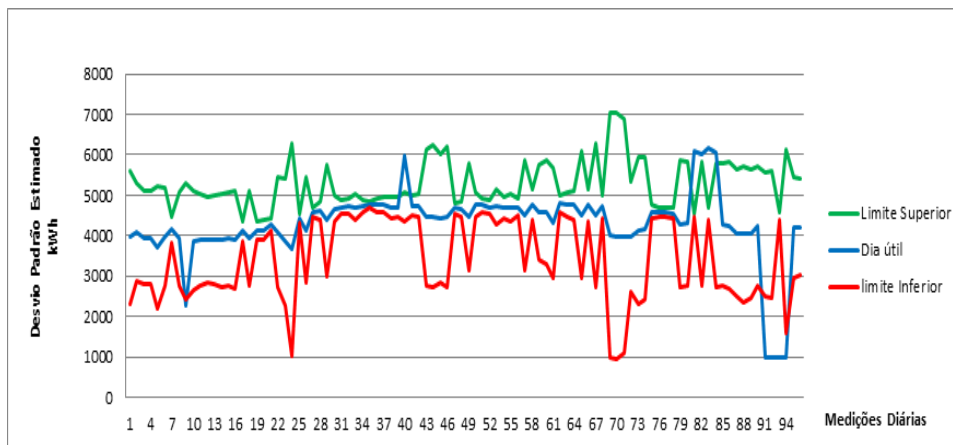


Figura 17 – Intervalo de confiança (96 medições de um dia útil) com “outlier”
 Fonte: o autor

Na análise dos dados discrepantes ou faltantes concluiu-se que estes podem ser relevantes para caracterizar o perfil do consumidor, então, o critério de decisão adotado foi o de não eliminação destes dados, tratando-os como dados faltantes “missing data”, tornando-os aceitáveis na série histórica de medições.

O diagrama de blocos da figura 18 resume basicamente a rotina do tratamento de “outlier”.



Figura 18 – Diagrama de blocos rotina tratamento de “outlier”
Fonte o autor

5.1.4. Tratamento de dados faltantes “missing data”

É muito comum ter observações com valores faltantes para uma ou mais características de entrada, sendo a abordagem usual para solucionar este problema a imputação (preenchimento) dos valores faltantes.

Uma maneira de lidar com o problema é determinar se os dados faltantes distorcem os dados observados. Grosso modo, os dados faltam de forma aleatória se o resultando da sua omissão independe do valor não observado [22].

Uma definição mais precisa é dada em Little e Rubin (2002). Suponhamos que y é o vetor de resposta e X é a matriz de entradas $N \times p$ com alguns dados faltantes. Denote por X_{obs} as entradas observadas em X e $Z = (y, X)$; $Z_{obs} = (y, X_{obs})$. Finalmente, se R é uma matriz com entrada ij se x_{ij} está faltando, os dados são ditos ausentes de forma aleatória (MAR) se a distribuição de R depende dos dados de Z somente através Z_{obs} , equação 3:

$$\Pr(R|Z, \theta) = \Pr(R|Z_{obs}, \theta) \quad (3)$$

Onde θ são quaisquer parâmetros na distribuição de R . Os dados são ditos faltantes completamente ao acaso (MCAR) se a distribuição de R não depende da observação ou dos dados faltantes, equação 4:

$$\Pr(R|Z, \theta) = \Pr(R|\theta) \quad (4)$$

MCAR é uma forte suposição de MAR, a maioria dos métodos de imputação confia em MCAR para validar-se.

Assumindo as características de dados faltantes aleatórios, existem várias maneiras de se proceder:

a) Descartar as observações com valores faltantes, se a quantidade relativa de dados faltantes for pequena;

b) Utilizar um algoritmo para lidar com os valores faltantes, generalizando uma modelagem aditiva, as observações faltantes com uma dada característica de entrada são omitidas quando os resíduos parciais são suavizados no algoritmo de ajuste, e os seus valores ajustados são definidos para zero. Uma vez que o curvas ajustadas tem a média zero, quando o modelo inclui uma intercepção, equivale a atribuir um valor médio ajustado para as observações faltantes;

c) Imputar todos os valores faltantes, na maioria dos métodos a abordagem de imputação é necessária. A tática mais simples é imputar o valor faltante com a média ou a mediana dos valores não faltantes de determinada função.

Após a imputação, os valores faltantes são normalmente tratados como se fossem observados. Isso ignora a incerteza devido à imputação, introduzindo incertezas adicionais em estimativas e previsões a partir do modelo de resposta. Pode-se medir essa incerteza adicional fazendo a imputação múltipla e, criando muitos conjuntos de treinamento diferentes [22].

Grandes massas de armazenamento de dados, como a das de medição de energia elétrica, normalmente possuem a característica de muitas lacunas durante as medições, que precisam ser tratadas, pois os dados faltantes podem interferir na caracterização do perfil dos consumidores.

Para o tratamento dos dados faltantes ou “*missing data*” da massa de dados de medições estudada nesta dissertação, foi utilizado o método estatístico de imputação múltipla por *substituição pela média*, que limpa a série de dados sem comprometer as características reais de consumo do cliente, aproximando as medições para os valores reais. Assim, detecta-se e corrige-se “*missings data*” (ausência de dados).

Desenvolveu-se então, uma rotina computacional de “*missing data*”, implementada no sistema SIMCAR, para onde foram importados os dados de medição dos clientes, composta de quatro passos:

- 1) O sistema busca no histórico de medições do cliente, um dia sem dados faltantes, similar ao dia com dados faltantes, ou seja, em que o comportamento de consumo tenha sido semelhante.

Ressalta-se, que para os feriados nacionais a regra estipulada é a mesma, porém nos demais feriados (estaduais e municipais), o tratamento é realizado com base em um feriado que ocorreu se possível, no mesmo dia da semana do feriado em questão. Caso não seja localizado, o sistema realizará uma nova busca, para identificar um dia onde o comportamento tenha sido o mais próximo possível do feriado em questão.

- 2) São determinados os limites inferiores (I_i) e superiores (I_s) do intervalo de medição faltantes, ou seja, o limite inferior será a medição realizada no horário anterior a medição faltante, e, o superior a primeira medição do horário da regularização da medição.

Na figura 19 é possível visualizar estes limites, onde às 12:30h ocorreu uma falha de medição, retornando ao normal às 14:30h, logo neste exemplo, o limite inferior será a medição de número 50 e o limite superior a medição de número 58, dentro do ciclo diário das 96 medições diárias.

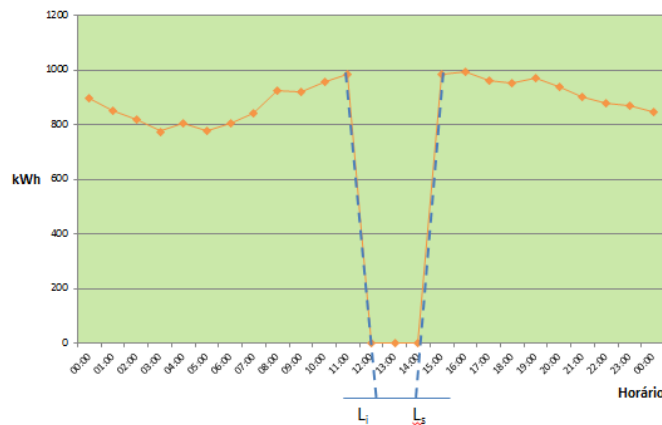


Figura 19 – Carga diária de um consumidor com “*missing data*”

Fonte: o autor

- 3) São realizadas as médias do dia com e sem “*missing data*”, equações 5 e 6. Aplicando estes resultados na equação 7,

encontra-se o dia médio característico do consumidor com os dados faltantes tratados.

Sendo:

DM_t – Dia com missing data

D_t - Dia sem missing data

l_i - Limite inferior do intervalo com missing data

l_s - Limite superior do intervalo com missing data

$$\bar{D} = \sum_{\substack{t=1 \\ t \neq (l_i; l_s)}}^{96} \frac{D_t}{96 - (l_s - l_i)} \quad (5)$$

$$\overline{DM} = \sum_{\substack{t=1 \\ t \neq (l_i; l_s)}}^{96} \frac{DM_t}{96 - (l_s - l_i)} \quad (6)$$

$$DM_k = \frac{D_k}{\bar{D}} \times \overline{DM} \quad (7)$$

$$k = l_i, \dots, l_s$$

- 4) Substituição dos dados faltantes “missing data”, obtendo-se assim, a curva carga com as medições faltantes tratadas, sem comprometer o perfil do cliente, conforme se observa na figura 20.

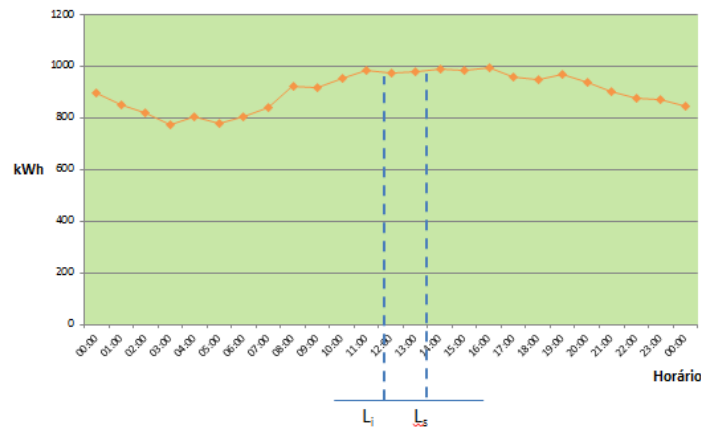


Figura 20 – Tratamento “missing data”

Fonte: o autor

5.1.5. Descrição do algoritmo de detecção

O algoritmo de detecção trabalha com as medições diárias dos consumidores de energia elétrica, ou seja, é alimentado com os horários de consumo, ponta e fora ponta e com as medições diárias a cada 15 minutos. Logo, do total das 96 medições diárias, 84 são no horário fora ponta e 12 são no horário de ponta.

O algoritmo foi construído seguindo três passos:

- 1) Cálculo das médias de demanda na ponta e fora ponta, formuladas matematicamente nas equações 8 e 9. A *média de demanda na ponta* representa o somatório da razão do consumo na ponta pelas medições na ponta; e a *média da demanda fora ponta*, o somatório da razão do consumo fora ponta pelas medições fora ponta.

Sendo:

CP_{id} – Consumo Ponta

CF_{id} – Consumo Fora da Ponta

i = medições diárias de 15 em 15 minutos; $i = 1, 2, \dots, 96$.

μ_{CP} – Média Horária Demanda Ponta

μ_{CF} – Média Horária Demanda Fora Ponta

l = início do horário na ponta

$$\mu_{CP} = \frac{\sum_{i=l}^{l+12} CP_{id}}{12} \quad (8)$$

$$\mu_{CF} = \frac{\sum_{\substack{i=l \\ i \neq l, l+12}}^{96} CP_{id}}{84} \quad (9)$$

- 2) Cálculo da estimativa de consumo, que neste algoritmo é representada em sua formulação matemática por (E), equação 10.

$$E = \frac{(\mu_{CP} - \mu_{CF})}{\mu_{CP}} \times 100 \quad (10)$$

A estimativa (E) pode assumir valores positivos ou negativos, sendo que se o resultado desta estimativa for positivo ($E > 0$), indicará que o consumidor possui características de consumo no horário de ponta, portanto, tem o perfil para aplicação de tecnologias alternativas de eficiência energética.

Portanto, têm-se as seguintes definições de respostas do algoritmo:

Cliente potencial – cliente que possui um consumo de energia praticamente constante, ou seja, com poucas variações no decorrer do dia, mantendo-se assim no horário de ponta.

Cliente muito potencial – cliente que consome muita energia no horário de ponta e pouca nos horários da madrugada.

Cliente que já modula – cliente que não apresenta consumo ou este é bem reduzido no horário de ponta.

- 3) O critério de decisão adotado no algoritmo é realizado com base no cálculo da estimativa de consumo (E). Se o valor da estimativa for positivo indica que o cliente tem potencial para utilização de tecnologia alternativa, caso contrário, se for negativa, sugere que o cliente já modula sua curva de carga, ou seja, possui algum sistema alternativo no horário de ponta do sistema elétrico.

O resultado positivo da estimativa (E) sugere que o cliente é potencial, mas sua potencialidade pode ser maior ou menor, de acordo com sua característica de uso da energia, variando de cliente para cliente, em função, da demanda solicitada, da forma e do tempo de utilização da energia.

Consideraram-se então no algoritmo duas possibilidades para o potencial dos clientes: *potencial e muito potencial*, e determinou-se um valor limiar (ϵ) para embasar a decisão entre estas condições.

Este valor limiar (ϵ) é um hiperparâmetro do algoritmo, exogenamente fixado em 10% nesta dissertação, percentual a partir do qual foi considerado valer o investimento em um sistema de GLD, podendo este ser variado de acordo com os interesses do projeto.

Os parâmetros utilizados para determinação deste limiar foram: a forma de utilização de energia do cliente, a avaliação do tempo de retorno do investimento em sistemas de eficiência energética e sua vida útil.

Analisando as curvas de cargas de vários clientes da concessionária, observou-se que alguns consumidores, apesar de possuírem um pequeno consumo no início do horário de ponta não seriam indicados para instalar uma tecnologia de eficiência energética, pois seu retorno ocorreria em um prazo

muito longo para reaver o investimento necessário em uma tecnologia desse porte, nestes casos algumas ações menores de eficiência energética já seriam suficientes para num menor tempo e custo, obterem retorno.

Nos indicadores econômicos de investimento para implantação da termoacumulação do estudo de caso realizado no Edifício da Bolsa de Valores do Rio de Janeiro (EBVRJ), anexo I, verifica-se que os investimentos nesta tecnologia são altos, apesar de ser atrativo e apresentar um retorno a um prazo razoavelmente curto, para o porte do cliente avaliado.

Estudos em sustentabilidade demonstram, que assim como a termoacumulação, geralmente investimentos em conservação de energia são relativamente altos, sendo o retorno ao longo do tempo favorável, havendo, porém, a necessidade de avaliação técnica e econômica de cada caso em particular, para se estimar este o tempo e comprando com a vida útil da instalação.

Portando o critério de decisão do algoritmo ficou estabelecido da seguinte forma:

Se:

$E \leq \epsilon$; ($\epsilon = 10\%$) : Cliente Potencial

$E > \epsilon$: Cliente Muito Potencial

$E < 0$: Cliente já modula na ponta

Desta forma, se o resultado dos cálculos da estimativa (E) for menor ou igual ao valor limiar, o cliente é identificado como potencial ao uso do sistema de termoacumulação, pois este resultado indicará que o mesmo consome energia no horário de ponta e durante a madrugada.

Contudo, se o resultado for maior que o valor limiar, o cliente é apontado como muito potencial, demonstrando que o mesmo consome pouca energia nos horários de madrugada e muita energia no horário de ponta.

E se o resultado da estimativa for menor que zero, indica que o cliente já modula na ponta, ou seja, possui algum sistema alternativo de GLD.

5.1.6. Fluxograma básico do algoritmo de detecção

A figura 21 representa o processamento básico do algoritmo de detecção de clientes potenciais ao uso de tecnologia de conservação de energia.

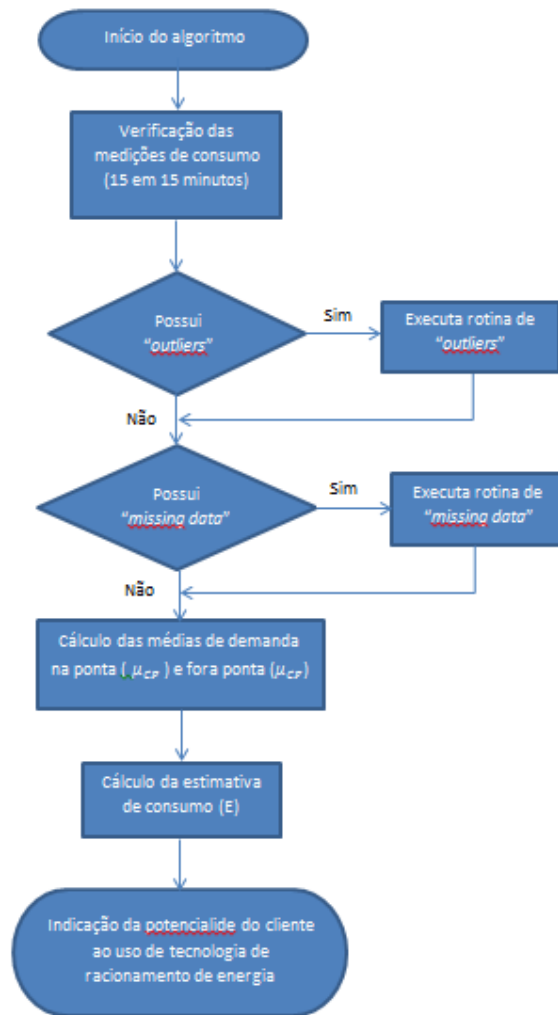


Figura 21 – Fluxograma básico do algoritmo de detecção
Fonte o autor

5.2. Sistema gerador de curvas de carga (SIMCAR)

O SIMCAR é um sistema computacional desenvolvido para a implementação do algoritmo de detecção e conseqüentemente o mapeamento das curvas de carga dos clientes das concessionárias, a partir dos dados de medição.

O sistema é voltado para a racionalização do uso de energia, possuindo interface "amigável", figura 22. Nele encontram-se também implementadas as rotinas de limpeza de dados: "outliers" e "missing data".

Para esta dissertação, realizada em consonância com o projeto de P&D de temoacumulação da LIGHT-ANEEL, o sistema foi direcionado para a

identificação de consumidores potenciais ao uso da termoacumulação (Manual de utilização Anexo II).

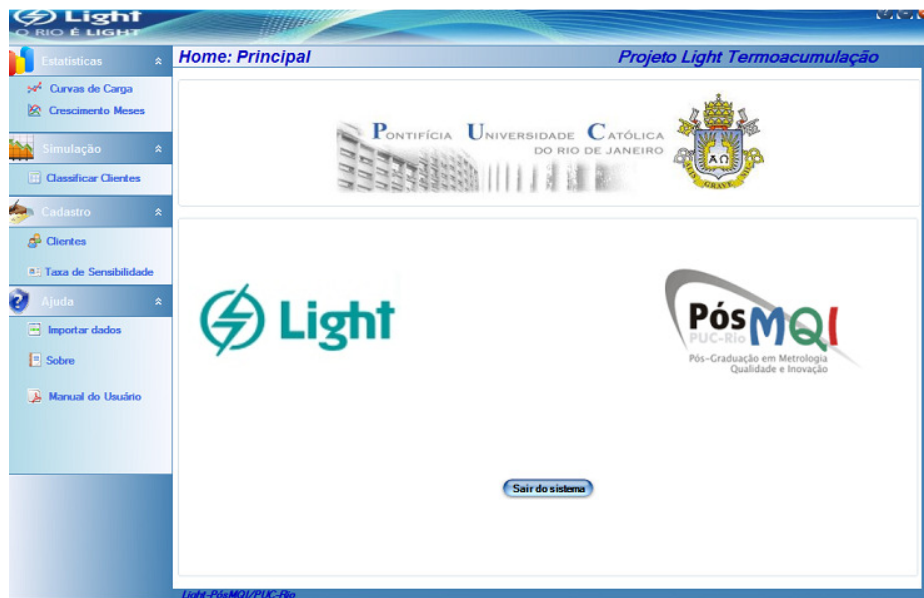


Figura 22 – Tela principal do sistema computacional
Fonte SIMCAR

5.2.1. Características do sistema

- Linguagem de programação Visual Basic.Net 2005.
- Base de dados SQL Server 2005.
- Módulos do sistema: Estatísticas, Simulação, Cadastro e Ajuda.

Input do sistema:

- Medições de demanda de energia com intervalos de quinze em quinze minutos.

Output do sistema:

- Mapeamento das curvas de carga diárias, mensais e anuais das medições a cada quinze minutos.
- Quatro curvas diárias com características diferenciadas de consumo: dias úteis, sábados, domingos e segundas feiras.
- Gráficos e tabelas.
- Importação de dados de planilhas com memória de massa em Excel.
- Classificação do potencial do cliente para a utilização ou indicação se já utiliza alternativa tecnológica no horário de ponta.
- Apresenta o comparativo do crescimento mensal de consumo das unidades.
- Apresenta o cadastro do cliente com os seus dados gerais, bem como o tipo de tarifa, CNAE, consumo mensal na ponta

- e fora de ponta, demanda na ponta e convencional, demanda ultrapassada na ponta e fora de ponta.
- Simula o comparativo do cliente potencial ao uso de tecnologia alternativa no horário de ponta em sua atual situação e utilizando a termoacumulação, bem como disponibiliza a projeção para o aumento ou redução de consumo na ponta e fora de ponta.

Notas:

Os dias úteis (DU) são os dias da semana de terça feira a sexta feira. A segunda feira no segmento comercial costuma ter um perfil de consumo diferenciado dos outros dias úteis da semana, por isso foi agrupada e tratada diferencialmente aos dias úteis comuns.

Na importação da memória de dados da planilha em Excel para o sistema, o mesmo verifica se a instalação já consta armazenada e caso esteja permite a atualização das informações.