

30/06/2013

CLASSIFICAÇÃO DE CLIENTES, ATENDIDOS PELA MT OU AT, ATRAVÉS DE ESTIMATIVAS DE SUAS CURVAS DE CARGA HORÁRIAS

Rodrigo Gonçalves Novais



CLASSIFICAÇÃO DE CLIENTES, ATENDIDOS PELA MT OU AT, ATRAVÉS DE ESTIMATIVAS DE SUAS CURVAS DE CARGA HORÁRIAS

Aluno: Rodrigo Gonçalves Novais

Orientador: Reinaldo Castro Souza

Coorientador: Maurício Frota

Trabalho apresentado com requisito parcial à conclusão do curso de Engenharia de Controle e Automação na Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil



Agradecimentos

Primeiramente, à minha família, principalmente à minha mãe, pelo apoio, carinho e compreensão demonstrados ao longo de minha vida acadêmica.

Ao meu grande orientador PhD. Reinaldo Castro Souza, sem o qual não conseguiria produzir este trabalho em tão pouco tempo.

Ao meu coorientador PhD. Maurício Nogueira Frota, por seu acolhimento e generosidade.

À Fernanda Particelli e à equipe de TI da Light, por ter disponibilizado todas as informações necessárias para o desenvolvimento deste trabalho que desenvolveu-se no âmbito do Projeto de P&D Light-Aneel Ref. 80-11-2.

Resumo

Este Projeto de Fim de Curso desenvolve-se no âmbito do Projeto de P&D Light-Aneel Ref. 80-11-2 que, essencialmente, denota uma ação voltada aos Grandes Clientes da Light (clientes conectados na Média Tensão). Esta monografia teve como **objetivo** o desenvolvimento e validação de um algoritmo para identificação de clientes potenciais ao uso de painéis fotovoltaicos. Esta necessidade é promovida pela tendência de aumento de consumo nos grandes centros urbanos e industriais do Rio de Janeiro, principalmente nos horários de ponta, impulsionados pelo crescimento da classe média e da economia. A **motivação** deste trabalho, atendendo aos interesses do projeto, é oferecer uma ferramenta de auxílio à decisão para o planejamento energético das concessionárias. Como **resultado**, o algoritmo desenvolvido foi capaz de classificar os consumidores de média, e alta, tensões, utilizando-se de dados fornecidos pelo TI da Light, em duas categorias: potencial ou não potencial (à implantação de painéis fotovoltaicos em sua rede). Dentre as **conclusões**, o algoritmo se mostrou eficaz na redução dos níveis de ponta de seus clientes, entretanto, o estudo de viabilidade econômica do projeto demonstrou-o inviável, uma vez que o tempo de retorno é demasiado longo.

Palavras-chave: Painéis fotovoltaicos; Energia solar; Consumo de energia; Eficiência energética; Horário de ponta.

HIGH AND MEDIUM VOLTAGE CLIENTS CLASSIFICATION BASED ON THEIR ESTIMATED HOURLY LOAD PROFILE

Abstract

This work **aims** to develop and validate an algorithm to identify potential photovoltaic panel users among a database of customers. This need is influenced by the trend of increasing consumption in the large urban and industrial centers of Rio de Janeiro, especially at peak hours, driven by the growing middle class and the economy. This work's **motivation**, serving the interests of the project, is to offer a tool to aid decision making for energy utilities and infrastructure planning. As a **result**, the proposed algorithm was able to classify medium, and high, voltage consumers, using data provided by the IT team at Light, into two categories: 'potential' or 'not potential' (to the deployment of photovoltaic panels on their grid). Among the **conclusions**, the algorithm was effective in reducing levels of consumption at customer's peak hours; however, the economic feasibility study of the project deemed it to be impractical, since the return on investment is too small.

Keywords: Photovoltaic panels; Solar Energy; Energy consumption; Energetic efficiency; Peak Consumption.

Sumário

1.	Introdução	1
1.1.	Tópicos relevantes	1
1.1.1.	Migração da Modalidade tarifária do cliente	1
1.1.2.	Definição das características do cliente beneficiado.....	2
1.1.3.	Viabilização econômica de investimento	2
1.1.4.	Verificação da eficácia da medida	2
1.2.	Verificação da eficácia da medida	2
2.	Modalidades Tarifárias	4
3.	Tratamento de dados	8
3.1.	Dados.....	8
3.1.1.	Estrutura.....	8
3.1.2.	Dimensões	8
3.1.3.	Tratamento	8
3.2.	Dados de radiação Solar	9
3.2.1.	Estrutura e dimensões	9
3.2.2.	Tratamento	9
3.3.	Painel escolhido	11
4.	Algoritmo de perfil	13
4.1.	Introdução	13
4.2.	Normalização.....	13
4.3.	Regressões Múltiplas	15
4.3.1.	Definição de variáveis	15
4.3.2.	Definição do Modelo	16
4.3.3.	Interpretação do modelo	18
4.3.4.	Avaliação do modelo.....	19
4.3.5.	Intervalos de confiança e classificação	22
5.	Análise de investimento	24
5.1.	Análise de Custos e Eficácia.....	24
5.2.	Análise de Pareto	25
5.3.	Retorno de investimento	26
6.	Conclusões.....	27
6.1.	Conclusões	27

6.2.	Comentários Finais.....	27
7.	Bibliografia.....	28

1. Introdução

Durante a primavera e o verão, as estações mais quentes do ano, o consumo de energia na cidade do Rio de Janeiro se eleva consideravelmente, impondo sobrecarga no sistema de abastecimento das concessionárias de energia elétrica. O problema se agrava nos horários de pico, quando ocorre a maior demanda, períodos em que o sistema de distribuição se torna mais vulnerável a sobrecargas e falhas. Investimentos por parte da concessionária se fazem necessários para evitar que os riscos associados à sobrecarga possam resultar em acidentes ou interrupção do fornecimento.

Para amenizar este problema, desde 1985, o governo brasileiro institui anualmente o horário de verão como um mecanismo de proteção ao sistema de distribuição. Comprovadamente, essa medida permite o deslocamento de carga para períodos menos solicitados, assim reduzindo a carga nos horários de pico, em média, de 4% a 5%¹ [FONTE ANEEL]; da mesma forma, reduz-se o consumo em cerca de 1%. Entretanto, resultado do desenvolvimento econômico e industrial, adicionalmente ao crescimento da classe média, o País experimenta expressivo aumento da demanda a um nível que o mecanismo do horário de verão, por si só, já não é suficiente para assegurar a proteção total do sistema de fornecimento. Apenas nos últimos três anos, de 2011 a 2013, a Light vem, a cada primeiro trimestre destes anos, experimentando um aumento de demanda expressivo por parte de seus consumidores, em média, 1,8% ao ano² [FONTE LIGHT CONSUMO]. Portanto, vê-se a necessidade de um plano mais ativo no combate ao consumo excessivo nos horários de ponta.

Neste contexto, o GLD³ [FONTE DISSERTAÇÃO PATRICIA], Gerenciamento pelo Lado da Demanda, vem sendo usado desde a década de 70, após a crise do petróleo. Este empreendimento envolve custos menores e retornos em um período mais curto de tempo se comparado aos investimentos pesados em infraestrutura que seriam necessários e, portanto, é um método eficiente para redução do consumo de energia no curto prazo, contribuindo para a estruturação e planejamento da matriz energética do país.

Um programa GLD envolve planejamento, análise e implementação de atividades que influenciam o consumidor a mudar o comportamento de sua curva de carga. A implantação de cada alternativa pode resultar num uso eficiente dos recursos e reduzir os custos para a empresa elétrica e o consumidor.

Existem várias alternativas para o gerenciamento de demanda, dentre eles a geração de energia fotovoltaica, sistema que produz energia elétrica a partir da irradiação solar. Como será exibido mais adiante, esta alternativa se mostra eficaz na redução de consumo de energia nos horários de pico.

Esta monografia abordará a alternativa de cogeração de energia elétrica através da utilização de painéis solares fotovoltaicos, alternativa tecnológica que o Projeto de P&D pretende propor aos grandes clientes da Light cujas curvas de carga confrontadas com o espectro de radiação solar confirmem um potencial favorável à conversão fotovoltaica.

1.1. Tópicos relevantes

Durante este estudo, todas as variáveis que envolvem a aplicação deste sistema devem ser levadas em consideração. Dentre elas, podem-se destacar:

- Migração de modalidade tarifária do cliente (item 1.1.1).
- Definição das características do cliente beneficiado (item 1.1.2).
- Viabilização econômica de investimento (item 1.1.3).
- Verificação da eficácia da medida (item 1.1.4).

1.1.1. Migração da Modalidade tarifária do cliente

Resumidamente, existem dois tipos de classes de consumidores: uma criada para os de baixa tensão (BT) e outra para os de alta, ou média, tensão (AT, MT). Estes grupos serão vistos com mais detalhes no capítulo 2. Os clientes BT, como característica desta classificação, são, obrigatoriamente, sujeitos à modalidade tarifária convencional, sem opção pelas tarifas horosazonais verde ou azul. Outra característica destes clientes é o seu baixo nível de consumo, que praticamente os eliminam do estudo.

Ademais, estes clientes não possuem um mecanismo de medição de consumo contínuo, como os clientes de MT ou AT possuem, tornando a análise de sua curva de carga muito limitada. Portanto, a classe de clientes a qual esta monografia se dirige, é a classe MT/AT, que possui três opções de modalidade tarifária: convencional, horosazonal verde e a horosazonal azul (estudadas no cap. 2).

1.1.2. Definição das características do cliente beneficiado

Os consumidores beneficiados com a inserção de painéis fotovoltaicos em suas redes serão aqueles que apresentam alto nível de consumo nos horários de pico, nos períodos de verão e primavera. Isto se deve ao fato de que a maior parte da carga das concessionárias de energia se concentra nestes períodos do ano. Este tema será abordado no capítulo 4.

1.1.3. Viabilização econômica de investimento

O estudo de viabilidade de investimento aborda a economia gerada pelo empreendimento comparada ao seu custo, gerando a taxa de retorno do mesmo. Os dados de custo – por painel – são encontrados no último item do cap. 3, onde a escolha do painel solar é realizada. Os dados de economia de investimento e taxa de retorno são abordados no cap. 5.

Ademais, pode-se calcular também o custo monetário do investimento, que pode ser obtido através da comparação entre a taxa de retorno do mesmo e a taxa de retorno de um investimento financeiro equivalente; em outras palavras, efetuado com a mesma quantia de dinheiro. No caso desta monografia, a taxa de retorno será comparada com a taxa real de rendimentos da poupança (ver cap. 5).

1.1.4. Verificação da eficácia da medida

Em última instância, o objetivo do projeto é a redução dos níveis de pico, observados nas estações de primavera e verão, aos níveis de consumo de pico médios ao longo do ano. Neste sentido, o estudo verificará se os painéis fotovoltaicos, aliados a radiação solar, serão capazes de eliminar tais anomalias no sistema de abastecimento (cap. 5).

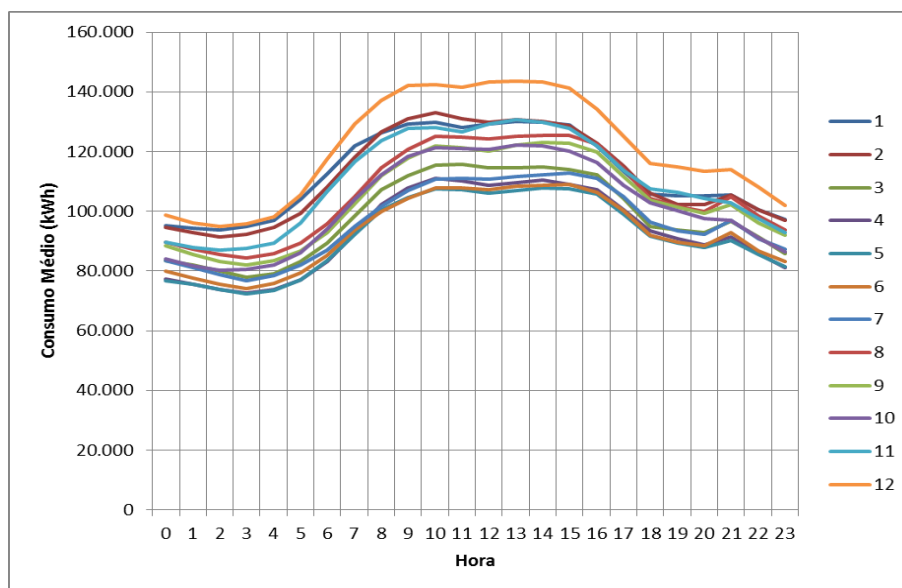
1.2. Verificação da eficácia da medida

Este Projeto de Fim de Curso desenvolve-se no âmbito do Projeto de P&D Light-Aneel Ref. 80-11-2 que, essencialmente, denota uma ação voltada aos Grandes Clientes da Light (clientes conectados na Média Tensão). O P&D propõe um modelo inovador de negócio e a melhoria dos padrões atuais de consumo, estimulando o uso de formas inovadoras de energia que permitem a transição de energia fóssil em renovável. Durante o seu desenvolvimento serão avaliadas três tecnologias de conversão fotovoltaica (filmes finos, mono e policristalino) objetivando aprimorar o processo de conversão fotovoltaica. Para tal implantará uma instalação piloto (cerca de 20 kW pico) no telhado do Museu de Energia da LIGHT assim permitindo o levantamento de dados, o acoplamento com a rede de energia da Light e a introdução de experimentos lúdico-educativos para elucidar aspectos conceituais da conversão fotovoltaica e o aproveitamento da energia para iluminar o espaço museológico.

No âmbito desse projeto, a Light, através de seus recursos de TI, forneceu dados valiosos para a estruturação deste estudo, fornecendo a massa de dados referente ao consumo horário de cerca de 2.000 clientes, ao longo de 2011. Com base nessa expressiva massa de dados, foi possível obter a curva de demanda real que a distribuidora atendeu durante todo o ano, bem como outras métricas críticas de interesse.

Por exemplo, pode-se exibir a média horária de cada mês do ano:

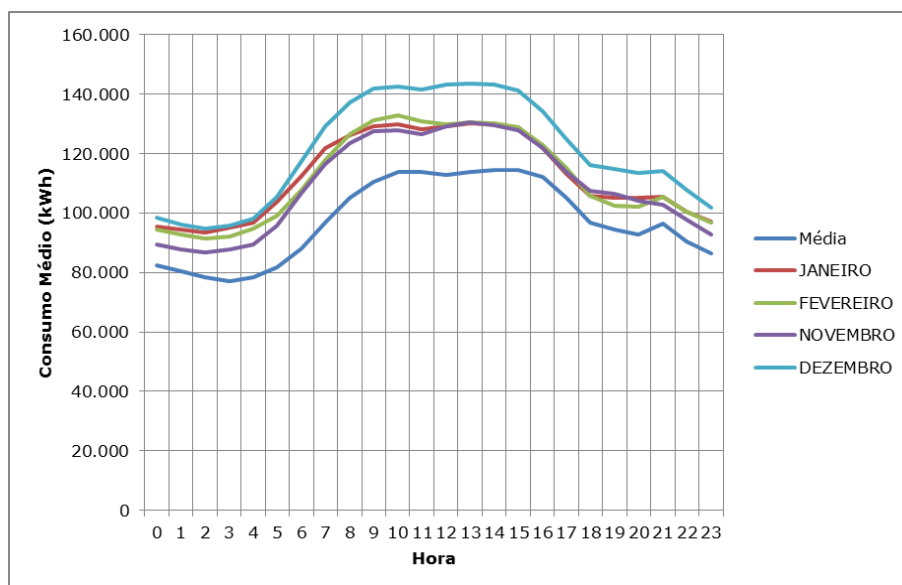
Figura 1 - Média horária de cada mês (ordenados numericamente)



De imediato, pode-se observar que a premissa de maior consumo durante as estações de verão e primavera é comprovada; os meses de Novembro, Dezembro, Janeiro e Fevereiro claramente se destacam ao serem comparados aos outros meses do ano. Obviamente trata-se de apenas uma amostra, mas já é o suficiente para que a robustez da metodologia sugerida seja posta à prova.

O objetivo proposto pelo estudo é reduzir as curvas destes meses à média anual composta pelos meses complementares, ou seja, os outros meses do ano. Desta forma:

Figura 2 - Médias horárias dos meses de interesse comparadas às médias do ano



Deseja-se que as curvas superiores do gráfico, exibido acima, sejam amortecidas para o mesmo patamar da curva da média. Isto geraria uma queda de energia suficientemente grande para que o sistema distribuidor deixe de estar sujeito a sobrecargas e, por conseguinte, a falhas catastróficas no fornecimento de energia regional.

2. Modalidades Tarifárias

Espera-se que, além da redução de consumo total do cliente, a aplicação de energia fotovoltaica possibilite a alteração da modalidade tarifária de fornecimento contratada. Esta oportunidade surge a partir da maior liberdade do cliente para modelar seu consumo, com o auxílio de produção de energia das células fotovoltaicas, às altas taxas impostas nos períodos de pico horários ou de secas anuais.

Clientes de baixa tensão, listados na tabela 1, são submetidos às taxas da tarifa convencional da Light⁴. **[FONTE LIGHT TABELAS]** Estas tarifas cobrem taxas de consumo padrão para qualquer horário do dia ou período do ano.

Tabela 1 - Classes de Consumidores de baixa tensão

Subgrupo	Classe
B1	Residencial
B1	Residencial baixa renda
B2	Rural
B2	Cooperativa de eletrificação rural
B2	Serviço público de irrigação
B3	Demais classes
B4	Iluminação pública

O enquadramento nesta categoria deve-se à demanda de tensão inferior a 2,3kV. [Light].

Para clientes de média tensão, listados na tabela 2, tem-se dois tipos de tarifas horo-sazonais oferecidas pela Light e, alternativamente, aqueles podem optar pela modalidade convencional de consumo. Para os clientes do grupo A2 (alta tensão) a definição de contrato dentre as normas da tarifa azul é compulsória.

Tabela 2 - Clientes de média e alta tensões

Subgrupo	Tensão
A2	88 kV a 138 kV (Alta tensão)
A3a	30 kV a 44 kV
A4	2,3 kV a 25 kV
AS* (Subterrâneo)	Menos de 2,3 kV

- Tarifa Verde: possui taxas de consumo variáveis que dependem da hora de ponta e períodos de seca/chuvas do ano. Vide tabela 3.

Tabela 3 - Tarifa horo-sazonal verde

ESTRUTURA HORO-SAZONAL VERDE						
Nível de tensão	Demanda	Demanda de Ultrapassagem	Consumo R\$/MWh			
			Ponta		Fora de Ponta	
	R\$/kW		Seca	Úmida	Seca	Úmida
A3a (30 a 44 kV)	15,82	31,64	1305,49	1276,81	186,95	170,31
A4 (2,3 a 25 kV)	16,70	33,40	1369,41	1340,73	186,95	170,31
AS (subterrâneo)	24.92	49.84	1425.66	1696.98	186.95	170.31

- Tarifa Azul: possui taxas de consumo e demanda variáveis que dependem da hora de ponta e períodos de seca/chuvas do ano. Vide tabela 4.

Tabela 4 - Tarifa horo-sazonal azul

ESTRUTURA HORO-SAZONAL AZUL								
Nível de tensão	Demanda R\$/kW		Demanda de Ultrapassagem R\$/kW		Consumo R\$/MWh			
					Ponta		Fora de Ponta	
	Ponta	Fora de ponta	Ponta	Fora de ponta	Seca	Úmida	Seca	Úmida
A2 (88 a 138 kV)	21,42	6,92	42,84	13,84	299,91	271,23	186,95	170,31
A3a (30 a 44 kV)	43,31	15,82	86,62	31,64	299,91	271,23	186,95	170,31
A4 (2,3 a 25 kV)	46,06	16,70	92,12	33,40	299,91	271,23	186,95	170,31
AS (subterrâneo)	50,95	24,90	101,90	49,84	299,91	271,23	186,95	170,31

- Tarifa convencional: possui taxas fixas de consumo e demanda, independente do horário ou período do ano de fornecimento. Vide tabela 5.

Tabela 5 - Tarifa convencional

ESTRUTURA CONVENCIONAL		
Nível de tensão	Demanda R\$/kW	Consumo R\$/MWh
A3a (30 a 44 kV)	47,00	189,72
A4 (2,3 a 25 kV)	49,86	189,72
AS (subterrâneo)	73,68	189,72

Vê-se que a tarifa de consumo nos horários de ponta da estrutura horo-sazonal verde é, em média, 350% mais cara do que a mesmas taxas pertencentes à estrutura azul. De forma semelhante, pode-se perceber, sem cálculo algum, que as taxas de consumo da estrutura convencional são, em média, muito inferiores as taxas variáveis. Como escolher, então, a taxa mais vantajosa para o perfil de consumo e demanda de um cliente?

Para responder esta pergunta, foram simuladas três diferentes demandas, com consumo variável, para a classe A3a de média tensão. As simulações levam em consideração que o perfil de consumo e demanda são constantes durante todo o mês da fatura. Não estão inclusos os ônus tributários.

Os três níveis de demanda modelados na página a seguir são: 300kW, 1000kW e 5000kW.

Figura 3 - Demanda a 1000kW

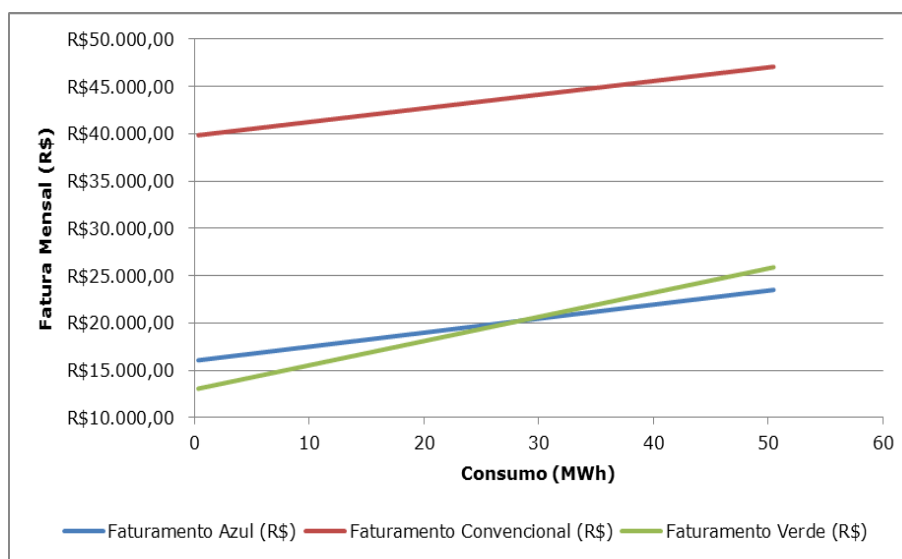


Figura 4 - Demanda a 300kW

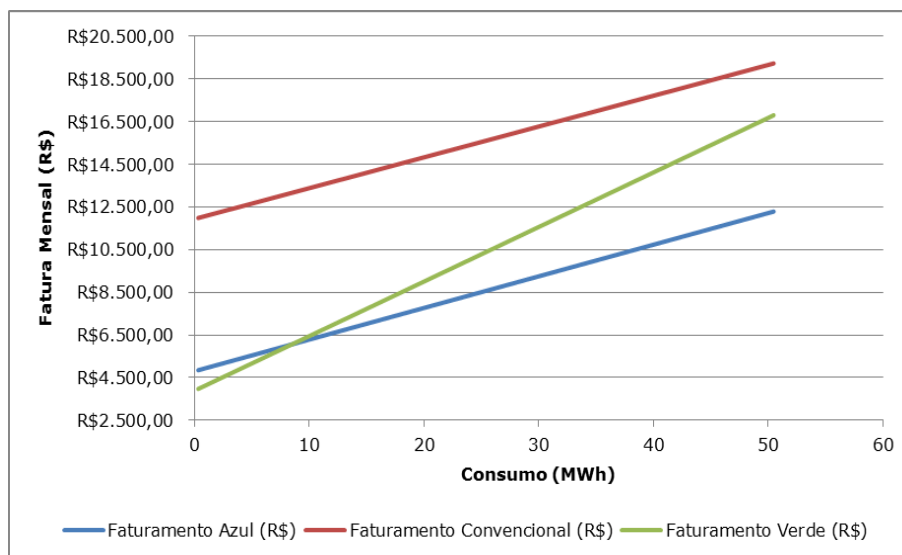
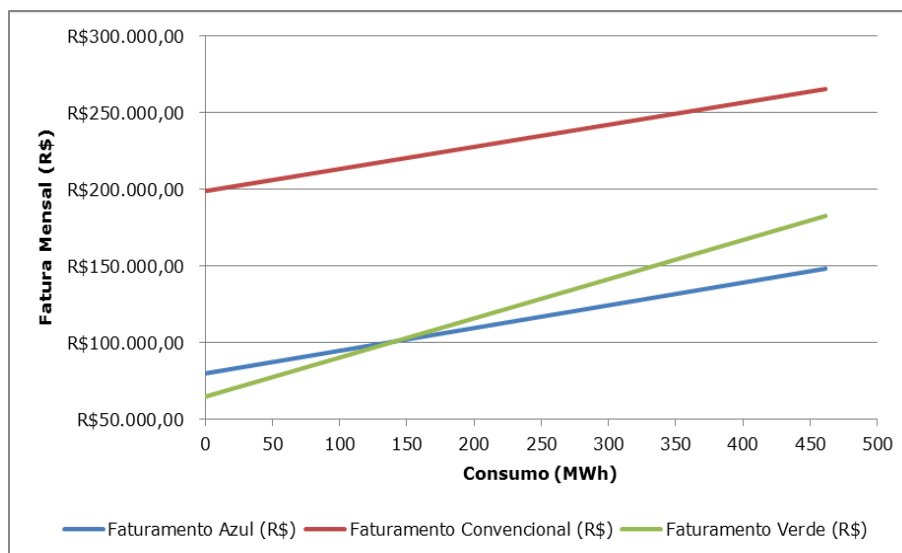


Figura 5 - Demanda a 5000kW



Através das simulações podemos perceber que há um ponto onde a taxa horo-sazonal verde cruza a modalidade tarifária azul, representando uma oportunidade de mudança no contrato tarifário do cliente. Vê-se, portanto, que a escolha entre modalidades tarifárias, a priori, pode sofrer alterações com a intervenção de um sistema alternativo de energia, que diminuiria a quantidade de consumo do cliente; por exemplo: se um cliente do tipo A3a, com demanda contratual de 1000W e consumo mensal de 50MWh obter uma redução de 30MWh consumo por mês com a instalação de um sistema fotovoltaico, seu consumo passa a ser 20MWh. Ao consultar-se a figura 1, é evidente que a atratividade da tarifa azul a um nível de consumo de 50MWh é mais alta do que todas as outras modalidades, entretanto, o mesmo não pode ser dito da faixa de consumo de 20MWh, onde a tarifa mais vantajosa passa a ser a verde.

3. Tratamento de dados

3.1. Dados

O foco do projeto é a escolha de usuários que mais possam desafogar a infraestrutura da Light com a adição de energia fotovoltaica em sua rede. Logo, vê-se a necessidade da obtenção de dados reais a fim de identificarmos o perfil mais adequado para este investimento.

Estes dados foram fornecidos pela Light e possuem as seguintes características constituintes:

3.1.1. Estrutura

A massa de dados fornecida é composta por 60 planilhas em formato CSV, contendo as seguintes colunas:

- QSUMC01 – Energia fornecida pela distribuidora (kWh).
- QSUMC02 – Energia reativa recebida do cliente (kVArh).
- Value – Descrição de campo não fornecida.
- Avail – Descrição de campo não fornecida.
- DT_Measur – Data de mensuração.
- ID_Meter_Read – Identificação de leitura.
- Counter – Descrição de campo não fornecida.
- ID_Meter – Identificação do medidor.
- Name – Nome em caracteres criptografados.
- Serial – Número de série do cliente.

Como visto acima, alguns atributos apresentam-se sem descrição uma vez que suas origens não nos foram fornecidas com devida clareza e, portanto, não serão utilizados neste estudo.

Quando a identificação do cliente temos 3 características a considerar: "Serial", "Name" e "ID_Meter". Como cada usuário pode possuir mais de um medidor, o campo utilizado para identificação dos mesmos será o "Serial". O único dado considerado de consumo será a primeira coluna, QSUMC01.

3.1.2. Dimensões

O conjunto de dados engloba o consumo de cerca de 2.200 clientes durante o ano de 2011; com um período de amostragem de 15 minutos. Isto se traduz em mais de 60 milhões de linhas, ou 24 Gigabits de dados. Todo o volume de dados será utilizado no projeto, sem fazer escolha de um cliente em especial, uma vez que procuramos o perfil mais adequado à utilização de energia fotovoltaica. Esta busca incita, necessariamente, a diferenciação entre os clientes potenciais dos demais, bem como a procura das razões para esta distinção.

Esta postura impõe certas dificuldades ao processamento de dados. Carregar um único arquivo contendo 60 milhões de linhas é impossível nos PCs atuais. As limitações, além de software, são físicas, uma vez que PCs raramente possuem mais do que 8 Gigabits de memória RAM. Sendo assim, não suportam carregar um arquivo desta magnitude. Portanto a utilização de um SGBD (Software gerenciador de banco de dados) fez-se necessária para a execução do estudo.

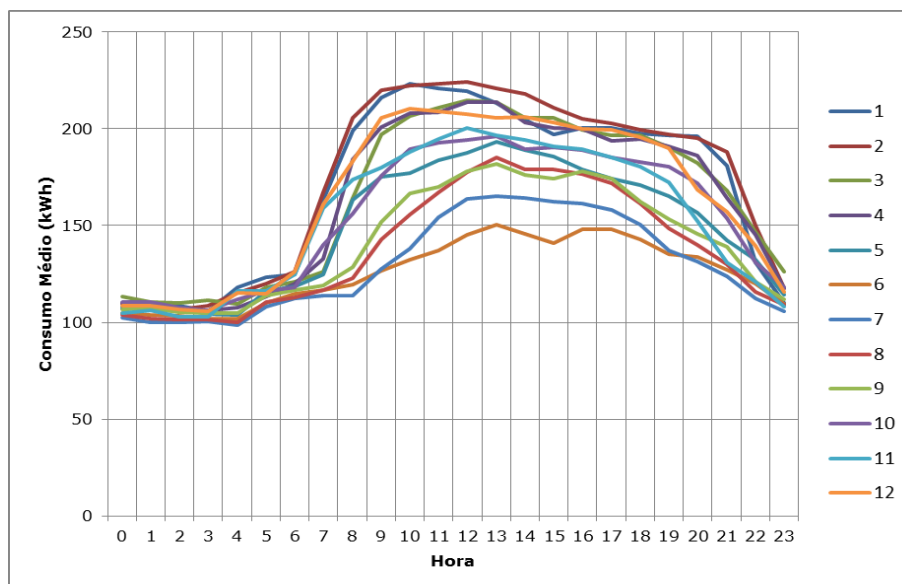
O programa escolhido foi o Microsoft SQL Server 2005®, fornecido a nós pelo departamento. Com este produto, fomos capazes de importar a massa de dados, trata-la e exporta-la para um software onde pudéssemos manuseá-los.

3.1.3. Tratamento

Dada à natureza da dimensão da massa de dados, a metodologia utilizada para reduzirmos a quantidade a ser processada foi a retirada da média horária mensal de cada cliente. Desta forma, compactamos o conjunto amostral inicial em 500 mil linhas, um número razoável de dados que agora podem ser tratados por outros softwares.

Em sua forma bruta, as 60 planilhas fornecidas não eram organizadas por clientes, tampouco no tempo. Portanto, a análise das curvas de carga dos clientes era impossível de ser realizada. Contudo, com a reestruturação dos dados, tornou-se possível a visualização destas características em sua forma média horária. Por exemplo:

Figura 6 - Média de consumo horária de um cliente



Donde se pode verificar, visualmente, o perfil de consumo de um cliente.

3.2. Dados de radiação Solar

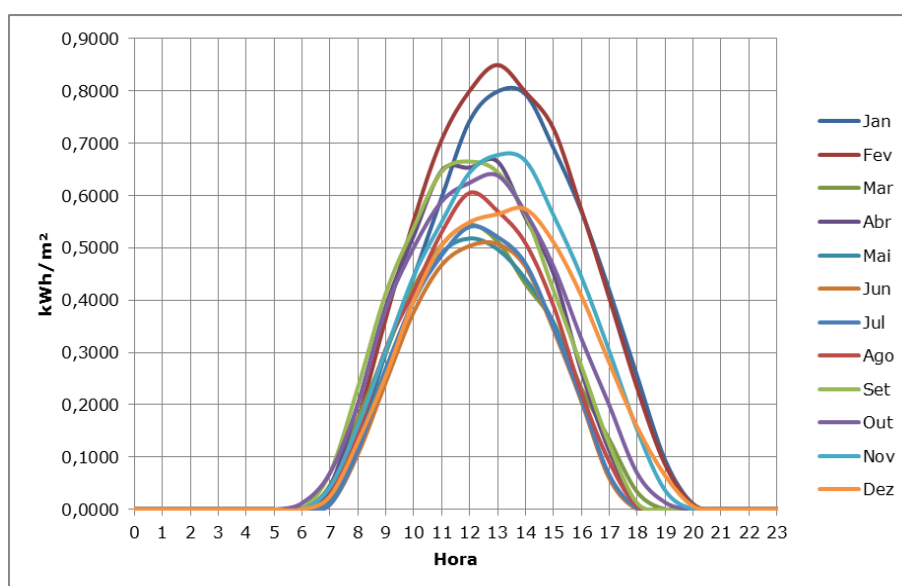
Para apurar a quantidade de energia que um painel solar é capaz de gerar no Rio de Janeiro (cidade), obtivemos os dados de irradiação solar do INMET (Instituto Nacional de Metrologia). Em seguida, devemos converter estes dados em energia produzida pelos painéis fotovoltaicos.

3.2.1. Estrutura e dimensões

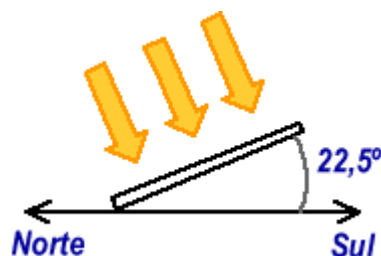
A estrutura dos dados de medição solar é muito simples: as colunas indicam as horas do dia (de 6h às 19h) e as linhas o dia do ano. Como no caso dos clientes, os dados são diários e envolvem todo o ano de 2011. Estas características resultam em uma planilha relativamente pequena, cujo tamanho é 178kbits, facilmente manuseável.

3.2.2. Tratamento

O tratamento inicial de dados é espelhado no aplicado nos dados dos clientes: médias horárias mensais. Desta forma, consegue-se extrair, compactamente, dados compatíveis com os da análise anterior. Pode-se, por exemplo, exibir um gráfico com as mesmas características do gráfico do cliente:

Figura 7 - Médias de irradiação solar horárias


Desta forma, podemos analisar a quantidade de energia economizada pelo cliente hora a hora. Contudo, a fim de converter-se, com razoável grau de precisão, a irradiação solar em energia gerada pelo painel, é necessário considerar a localização geográfica da cidade do Rio; mais especificamente, sua latitude em relação ao Equador. Como o grau de inclinação do sol no céu é modificado conforme andamos ao sul, ou norte, do equador, precisamos quantificar a alteração da energia captada pelo painel conforme sua inclinação muda em relação ao solo (fig. 5). As medidas realizadas pelo INMET foram horizontais em relação ao solo.

Figura 8 - Exemplo de inclinação de um painel solar (ângulo meramente ilustrativo)


O portal da CRESESB (Centro de Referência para Energia Solar e Eólica Sérgio de Salvo Brito) disponibiliza, em sua página 'SunData', dados da irradiação solar mensais (média diária mensal). Na mesma página, estão tabeladas algumas inclinações do sensor, juntamente com suas leituras de irradiação.

Por exemplo, temos as seguintes médias calculadas para o ano de 2013⁵:

Tabela 6 - Insolação diária média por mês por inclinação

Medições realizadas no Jardim Botânico						
Inclinação	Jan	Fev	Mar	Abr	Mai	Jun
0° N	5,86	5,67	5,22	4,06	3,83	3,22
23° N	5,29	5,41	5,38	4,56	4,76	4,14
21° N	5,36	5,45	5,39	4,54	4,7	4,08
43° N	4,83	5,07	5,23	4,62	5	4,41

Tabela 7 - Insolação diária média por mês por inclinação (continuação)

Medições realizadas no Jardim Botânico							
Jul	Ago	Set	Out	Nov	Dez	Média	Delta
3,78	4,06	4,22	4,86	5,28	5,61	4,64	2,64
4,85	4,77	4,47	4,74	4,85	5,01	4,85	1,27
4,78	4,73	4,47	4,77	4,91	5,08	4,85	1,37
5,15	4,9	4,41	4,5	4,47	4,55	4,76	0,82

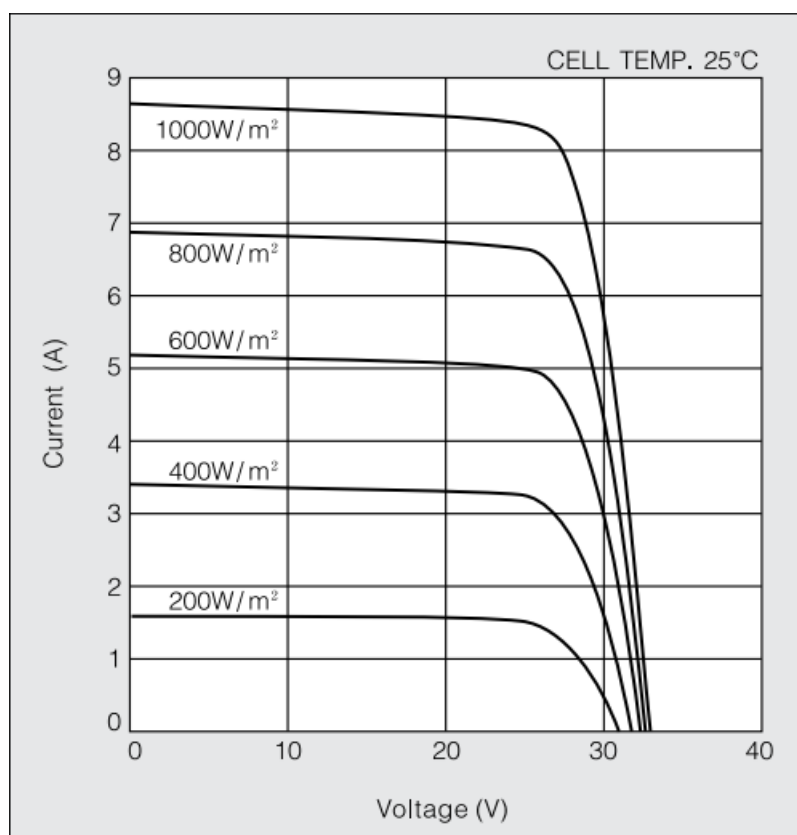
[FONTE CRESESB] Onde se pode perceber que a angulação com maior média de incidência diária são 21° e 23° ao Norte. Como a primeira inclinação é a que capta a maior média de irradiação solar, no verão e na primavera, os painéis instalados deverão permanecer planos em relação ao solo para melhor suavizar a curva de consumo durante o mesmo período.

Finalmente, aplica-se a taxa de eficiência do painel para obter-se uma estimativa de geração diária de energia. Este tema será abordado mais adiante.

3.3. Painel escolhido

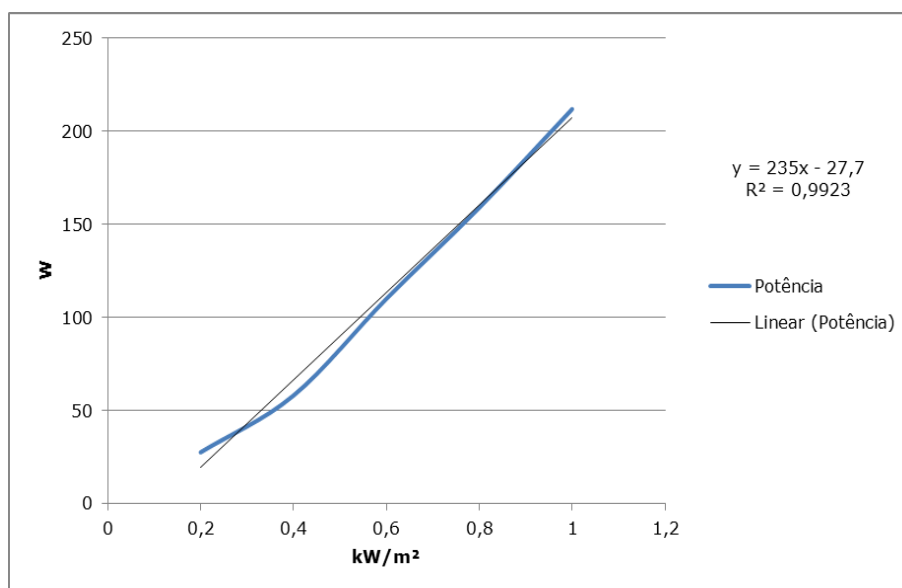
O painel escolhido foi o Kyocera KD210GX-LP, capaz de gerar 210W de potência @ 1000kW/m². No datasheet fornecido pelo fabricante não são encontrados taxas de conversão precisas, apenas uma declaração de que 'a eficiência de conversão [...] é superior a 16%'⁶. No entanto, o fabricante nos fornece as curvas de tensão-corrente do produto, em diferentes níveis de irradiação.

Figura 9 - Curvas de Potência do equipamento



Com estes dados, então, pode-se traçar uma curva (aproximada) que relaciona a potência gerada pelo painel com a energia irradiada pelo sol.

Figura 10 - Modelagem da geração do painel



Pode-se perceber que a resposta do sistema é altamente linear e é facilmente modelada por uma regressão linear (MQO, Mínimos Quadrados Ordinários).

PREÇO ESTIMADO: R\$ 2.223,96 (amazon.com)

4. Algoritmo de perfil

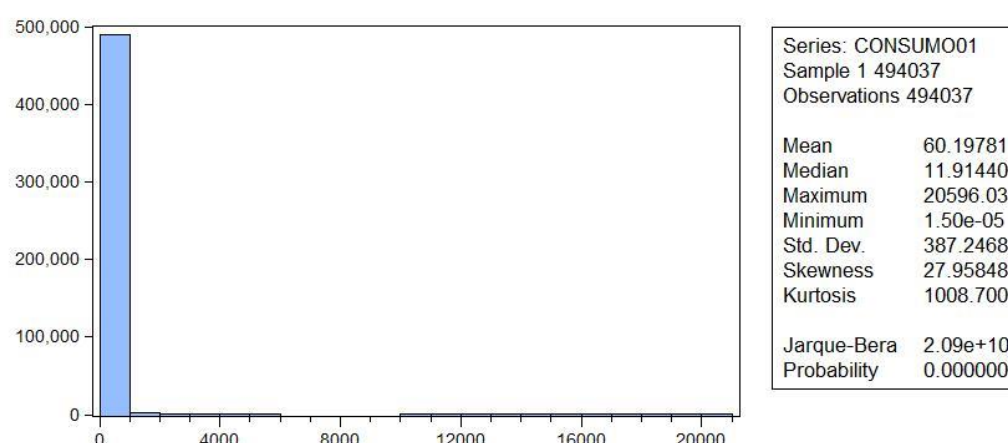
4.1. Introdução

A busca pelo perfil do cliente mais provável de se beneficiar com a adição de painéis fotovoltaicos em sua rede começa pela sua definição. Esta deriva das características do ciclo de iluminação solar bem como do perfil de consumo do cliente.

Como apontado no capítulo anterior, a radiação solar diária não é constante; pelo contrário, sofre variações ao longo do dia e, dependendo da época do ano, pode ser reduzida por quase 40%. Levando em consideração que a maior parte da energia irradiada pelo sol encontra-se no intervalo das 11h às 16h – fora deste período a energia solar nunca ultrapassa 50% de sua capacidade máxima – procura-se, nos dados dos clientes, um perfil de consumo energético também elevado durante este período. Desta forma, garante-se que a energia produzida pelos painéis será convertida em benefícios ao consumidor, atenuando sua curva de carga. Esta escolha é reforçada pela ausência de baterias no projeto, o que torna consumo de energia gerada uma necessidade imediata.

Tendo em vista a grande variabilidade de consumo presente entre clientes na massa de dados fornecida, o critério de comparação entre os mesmos cabe a outro indicador. A seguir, exibem-se o conjunto de parâmetros estatísticos descritivos da variável.

Figura 11 - Estatísticas descritivas do consumo



Portanto, a metodologia abordada por este projeto considera a normalização do consumo seu principal parâmetro de diferenciação para prospecção de potenciais clientes. Este indicador será abordado no tópico seguinte.

4.2. Normalização

Considere as seguintes definições de variáveis:

- \bar{h} é a média horária mensal da série de consumo de um cliente.
- $\max_{\text{mês}} h$ é a média horária máxima do mês.
- A normalização dos dados de consumo se dá da seguinte maneira:

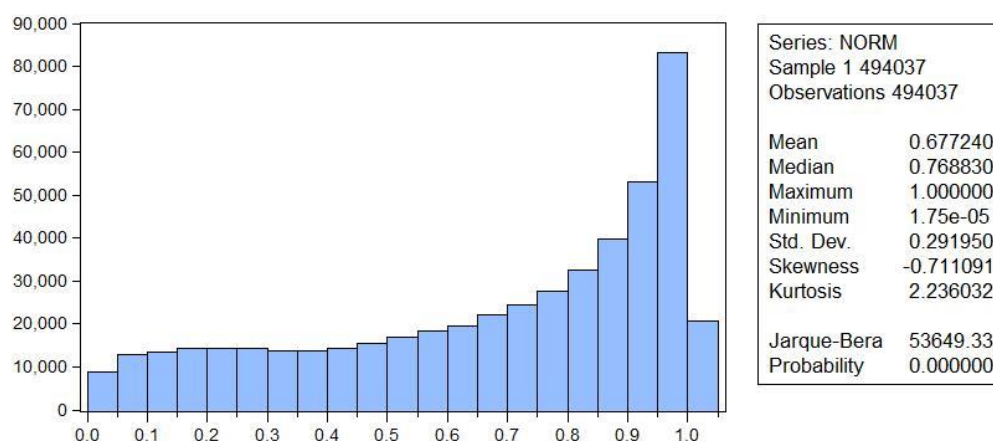
Equação 1 - Normalização pela média horária máxima mensal

$$p_i = \frac{\bar{h}}{\max_{\text{mês}} h}$$

Onde p_i representa a média horária normalizada pela máxima média horária do mês, para cada mês de cada cliente. Desta forma, o domínio de p equivale a: $\{p \mid x \in X\} = [0..1]$ onde x é um valor de consumo qualquer do conjunto amostral X fornecido pelo banco de dados.

Finalmente, obtém-se uma variável capaz de representar a variação diária de consumo de cada cliente sem que esta seja enviesada para os clientes que mais consomem. Agora, podem-se extrair algumas estatísticas descritivas da variável:

Figura 12 - Estatísticas descritivas do Consumo normalizado pela máxima mensal



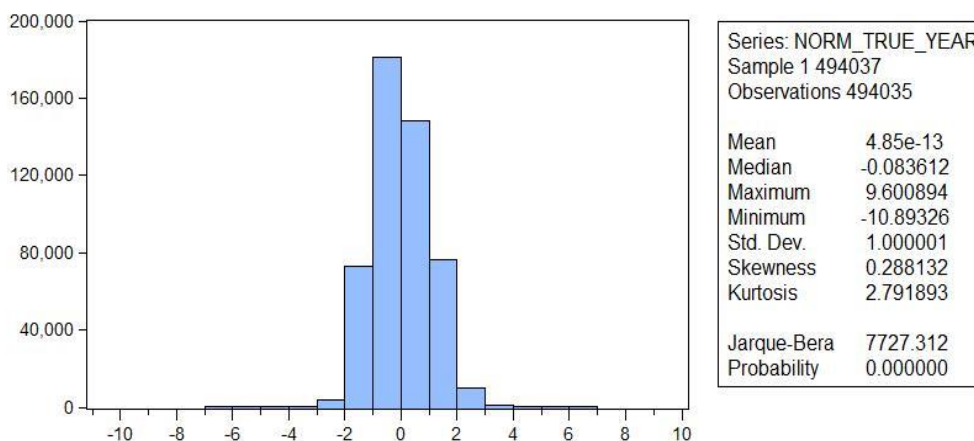
Alternativamente, cria-se um novo método de normalização, desta vez normalizando a variável na forma padrão, ou seja, esta passa a apresentar variância unitária e média nula. Para tanto, faz-se:

Equação 2 - Normalização padrão do consumo mensal

$$p_i = \frac{h_i - \bar{h}}{\sqrt{\frac{1}{n-1} \cdot \sum_{i=0}^n (h_i - \bar{h})^2}}$$

O que se espera, através desta transformação, é melhorar o teste de normalidade da variável. Desta forma, para o mesmo tipo de descrição estatística, obtemos:

Figura 13 - Estatísticas descritivas da normalização padrão do consumo



De fato, a melhoria em relação ao teste de normalidade anterior é visível, entretanto, a variável ainda não possui as características necessárias para ser classificada como uma variável normal. Entre estas disparidades, estão o alto valor de 'Skewness' (o terceiro momento da série) e a 'Kurtosis' (quarto momento da série) extremamente positiva, assemelhando a distribuição da amostra à de Laplace, leptokurtica. Portanto, como era de se esperar, o teste de Jarque-Bera rejeita a hipótese nula de normalidade da série.

4.3. Regressões Múltiplas

Para a classificação dos clientes, o algoritmo utiliza-se de regressões lineares múltiplas como sua principal ferramenta de diferenciação de perfis. A priori, definem-se as variáveis constitutivas do modelo e, a posteriori, uma análise detalhada do mesmo será realizada.

4.3.1. Definição de variáveis

A variável explicada, vastamente estudada no tópico anterior deste capítulo, o consumo normalizado, será utilizada no modelo proposto. As variáveis explicativas são binárias, chamadas 'dummies', e representam os períodos do dia e as estações do ano; com exceção de uma delas, o qual forma a base de comparação entre os diferentes períodos e estações (caso baseline). Como é natural deste tipo de variável, quando esta se encontra na classificação de interesse, assume o valor unitário; caso contrário, esta se torna nula.

Uma variável binária é utilizada no caso onde temos níveis qualitativos distintos a serem estudados que afetam o comportamento de nossa variável explicativa. No caso deste estudo, têm-se quatro níveis para cada período (estações do ano e períodos do dia), o que gera três variáveis classificativas para cada um. Isto ocorre, pois, uma delas se torna o caso baseline.

Portanto, podemos definir as seguintes variáveis binárias e suas variáveis auxiliares, mês e hora:

Mês: define-se m_i a variável que representa o mês relacionado à média horária p_i .

Verão = V :
$$\begin{cases} 1 & \text{se } m_i \in [1, 2, 3]. \\ 0 & \text{caso contrário.} \end{cases}$$

Outono = O :
$$\begin{cases} 1 & \text{se } m_i \in [4, 5, 6]. \\ 0 & \text{caso contrário.} \end{cases}$$

Primavera = P :
$$\begin{cases} 1 & \text{se } m_i \in [10, 11, 12]. \\ 0 & \text{caso contrário.} \end{cases}$$

Hora: define-se h_i a variável que representa a hora relacionada à média horária p_i .

Manha = M :
$$\begin{cases} 1 & \text{se } h_i \in [6, 7, 8, 9, 10, 11]. \\ 0 & \text{caso contrário.} \end{cases}$$

Tarde = T :
$$\begin{cases} 1 & \text{se } h_i \in [12, 13, 14, 15, 16, 17]. \\ 0 & \text{caso contrário.} \end{cases}$$

Noite = N :
$$\begin{cases} 1 & \text{se } h_i \in [18, 19, 20, 21, 22, 23]. \\ 0 & \text{caso contrário.} \end{cases}$$

Pode-se perceber que não existem variáveis para a madrugada e o inverno. Isto se deve ao fato deste ser, justamente, o caso 'baseline'.

4.3.2. Definição do Modelo

A priori, o modelo mais óbvio que se desprende da definição de variáveis é:

Equação 3 - Modelo Irrestrito (completo)

$$p_i = c + \beta_1 V_i + \beta_2 O_i + \beta_3 P_i + \beta_4 M_i + \beta_5 T_i + \beta_6 N_i + u$$

Onde β_n representa o coeficiente obtido através de MQO; c o coeficiente linear a ser estimado pelo mesmo método e u o resíduo da equação. Desta forma, obtemos:

Tabela 8 - Descrição estatística do modelo irrestrito utilizando a normalização mensal máxima

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.497569	0.000965	515.7555	0.0000
VERAO	0.005961	0.001073	5.554638	0.0000
OUTONO	0.003792	0.001043	3.634233	0.0003
PRIMAVERA	0.001462	0.001003	1.457348	0.1450
MANHA	0.309938	0.001045	296.6679	0.0000
TARDE	0.302435	0.001045	289.5289	0.0000
NOITE	0.095513	0.001045	91.40385	0.0000
R-squared	0.209388	Mean dependent var		0.677240
Adjusted R-squared	0.209379	S.D. dependent var		0.291950
S.E. of regression	0.259593	Akaike info criterion		0.140607
Sum squared resid	33291.83	Schwarz criterion		0.140765
Log likelihood	-34725.56	Hannan-Quinn criter.		0.140652
F-statistic	21806.79	Durbin-Watson stat		0.254875
Prob(F-statistic)	0.000000			

Primeiramente, podemos notar que as variáveis das estações do ano são as que mais chegam perto de serem rejeitadas. Primeiro porque seus coeficientes têm magnitudes duas ordens de grandeza inferiores se comparados às magnitudes dos coeficientes das variáveis relacionadas aos períodos do dia, e, ao mesmo tempo, possuem valores muito pequenos para o teste da distribuição T de Student. O coeficiente da variável 'Primavera' chega a ser estatisticamente rejeitado ao nível de significância $\alpha = 10\%$, quando o convencionalmente usado é de $\alpha = 5\%$. Isto se deve, principalmente, ao estilo de normalização adotado, que divide as observações pela máxima do mês, eliminando assim parte da sazonalidade anual da série.

Em outra simulação, adotou-se a normalização com base na máxima anual, melhorando assim os coeficientes do teste T de Student para estes regressores:

Tabela 9 - Estatística das variáveis mensais do modelo irrestrito normalizado anualmente

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VERAO	0.075710	0.001057	71.64712	0.0000
OUTONO	0.008684	0.001027	8.452849	0.0000
PRIMAVERA	0.038566	0.000987	39.05466	0.0000

Entretanto, o R^2 desta simulação ficou ainda inferior aos cerca de 21% encontrado na anterior, tornando o modelo original mais interessante ao estudo e aplicação ao algoritmo.

De volta à tabela original (adicionar legendas), pode-se perceber que, como mencionado acima, o R^2 do modelo ajustado é bastante baixo e, por conseguinte, seus regressores, apesar de serem relevantes e estatisticamente significantes (com exceção de Primavera), não explicam satisfatoriamente o efeito dos mesmos na variável explicada, a normalização do consumo. Neste contexto, vale lembrar que, no

tópico 'normalização', abordam-se duas metodologias distintas para o tratamento do consumo, a normalização pela máxima e a normalização padrão da variável. Portanto, é natural testar a outra transformação da variável aplicada ao mesmo modelo. Desta forma, obtêm-se:

Tabela 10 – Estatísticas Descritivas do modelo irrestrito utilizando a normalização padrão

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.754824	0.003133	-240.9524	0.0000
VERAO	-6.25E-05	0.003485	-0.017927	0.9857
OUTONO	-0.000213	0.003388	-0.062862	0.9499
PRIMAVERA	-9.82E-06	0.003257	-0.003014	0.9976
MANHA	1.266937	0.003392	373.4567	0.0000
TARDE	1.260914	0.003392	371.7345	0.0000
NOITE	0.490391	0.003393	144.5229	0.0000
R-squared	0.289483	Mean dependent var	-3.98E-14	
Adjusted R-squared	0.289475	S.D. dependent var	1.000001	
S.E. of regression	0.842927	Akaike info criterion	2.496143	
Sum squared resid	350997.4	Schwarz criterion	2.496300	
Log likelihood	-616544.0	Hannan-Quinn criter.	2.496187	
F-statistic	33544.51	Durbin-Watson stat	0.462300	
Prob(F-statistic)	0.000000			

De imediato, vê-se uma melhora no R^2 da regressão, o que faz deste um modelo melhor do que o anterior, no sentido de que explica melhor a interação (linear) entre os regressores e a variável explicada. De forma semelhante ao ocorrido no modelo anterior, os coeficientes das estações do ano tornam-se estatisticamente insignificantes e irrelevantes, fruto da normalização mensal dos dados. Se normalizado anualmente, o modelo passa a aceitar estes regressores, como no modelo anterior, mas, ao mesmo tempo, diminui-se o R^2 da regressão, tornando-a menos interessante para o dever classificativo que o modelo terá no algoritmo. A seguir, expõem-se apenas os regressores estimados para esta variável.

Tabela 11 - Estatística das variáveis mensais do modelo irrestrito normalizado da forma padrão

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VERAO	0.400946	0.003606	111.1906	0.0000
OUTONO	0.053757	0.003506	15.33475	0.0000
PRIMAVERA	0.196842	0.003370	58.41480	0.0000

Portanto, sem prejuízos para o entendimento do algoritmo e suas saídas produzidas, estas variáveis serão excluídas do modelo e deixarão de ser estudadas. Ao mesmo tempo, em consonância com a finalidade do projeto de diminuição da ponta de consumo nos meses de pico, restringe-se o conjunto amostral apenas aos meses de verão e primavera.

Finalmente o modelo escolhido é:

Tabela 12.1 - Estatísticas descritivas do modelo final

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.730251	0.003376	-216.3169	0.0000
MANHA	1.326706	0.004773	277.9555	0.0000
TARDE	1.184814	0.004772	248.2684	0.0000
NOITE	0.408314	0.004774	85.53042	0.0000
R-squared	0.299790	Mean dependent var		2.52E-13
Adjusted R-squared	0.299782	S.D. dependent var		1.000002
S.E. of regression	0.836792	Akaike info criterion		2.481535
Sum squared resid	172206.1	Schwarz criterion		2.481704
Log likelihood	-305144.1	Hannan-Quinn criter.		2.481584
F-statistic	35097.89	Durbin-Watson stat		0.466855
Prob(F-statistic)	0.000000			

Com um R^2 cerca de 10% maior do que o inicialmente explorado.

4.3.3. Interpretação do modelo

Primeiramente, vamos escrever a equação do modelo escolhido:

Equação 4 - Equação do modelo final

$$p_i = c + \beta_1 M_i + \beta_2 T_i + \beta_3 N_i + u$$

Sabemos também que, como normalizado na forma padrão, p_i representa o desvio padrão do consumo em relação à média de cada cliente. O intercepto c representa a média desta medida no caso 'baseline', enquanto que β_n representa a contribuição da média das observações classificadas por n à média do baseline. Esta interpretação, um tanto conveniente, pode ser deduzida da seguinte forma:

Toma-se o valor esperado da equação com todas as variáveis binárias nulas:

$$E(p_i | madrugada) = c + \beta_1 \cdot 0 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + (E(u | p_i))$$

$$\text{Assumindo que: } (E(u | p_i) = 0)$$

$$\text{Portanto: } E(p_i | madrugada) = c$$

Assim, tem-se que, de fato, c representa a média do caso baseline. Para β_n , o raciocínio é análogo. Assumindo que $n = M$:

$$E(p_i | manhã) = c + \beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_3 \cdot 0 + (E(u | p_i))$$

$$\text{Assumindo que: } (E(u | p_i) = 0)$$

$$\text{Portanto: } E(p_i | manhã) = c + \beta_1$$

Equação 5 – Equação final da prova da interpretação correta dos coeficientes da regressão

$$\text{Finalmente: } E(p_i | manhã) - E(p_i | madrugada) = \beta_1$$

Logo, β_1 nada mais é do que a diferença entre a média das madrugadas e a média das manhãs. Em outras palavras, os regressores estimados representam, em desvios padrão, a diferença entre as médias de cada uma das classificações ao caso baseline.

4.3.4. Avaliação do modelo

Por conveniência, repete-se a tabela anteriormente exibida neste capítulo:

Tabela 13 – Tabela 12 repetida

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.730251	0.003376	-216.3169	0.0000
MANHA	1.326706	0.004773	277.9555	0.0000
TARDE	1.184814	0.004772	248.2684	0.0000
NOITE	0.408314	0.004774	85.53042	0.0000
R-squared	0.299790	Mean dependent var	2.52E-13	
Adjusted R-squared	0.299782	S.D. dependent var	1.000002	
S.E. of regression	0.836792	Akaike info criterion	2.481535	
Sum squared resid	172206.1	Schwarz criterion	2.481704	
Log likelihood	-305144.1	Hannan-Quinn criter.	2.481584	
F-statistic	35097.89	Durbin-Watson stat	0.466855	
Prob(F-statistic)	0.000000			

A avaliação de um modelo de regressão é basicamente realizada através dos seguintes procedimentos:

- Diagnósticos: investiga se os pressupostos assumidos para os erros do modelo são observados (descorrelatados com a variável explicada).
- Medidas de aderência (goodness of fit): investigam a qualidade do ajuste do modelo aos dados.

Como o modelo será utilizado para fins de classificação e previsão, as medidas de aderência são genericamente mais importantes. Primeiramente, trata-se de fazer os diagnósticos do modelo.

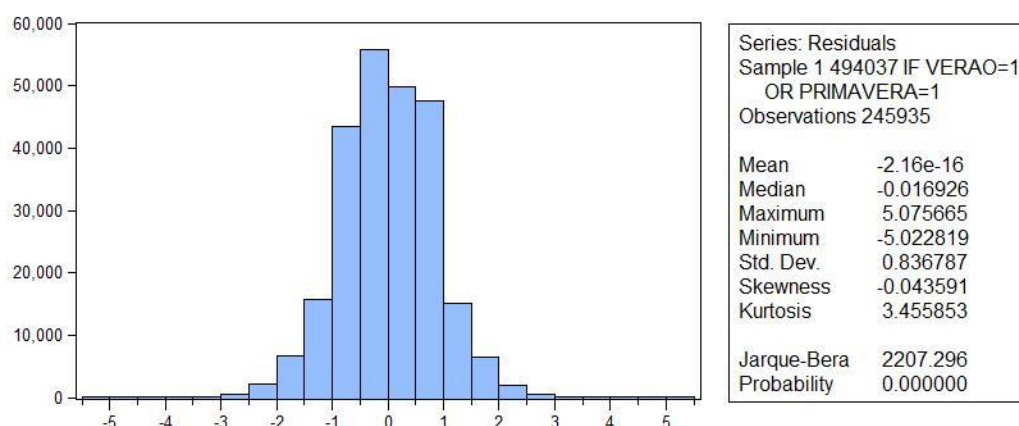
Para que os testes de hipótese do modelo tenham validade, o ruído da regressão precisa obedecer as seguintes regras:

- Deve possuir média nula: $E(u_i|p_i) = 0$
- Ter variância constante: $Var(u|p_i) = \sigma^2$
- Descorrelatado: $E(u_i u_j | p_i p_j) = 0$

Somente desta forma o Teorema de Gauss Markov afirma que os estimadores de MQO são os de menor variância (os mais eficientes) dentro da classe dos estimadores lineares e não enviesados. Isto é importante, uma vez que queremos estimar os perfis que mais se adequam ao projeto de maneira precisa.

A seguir, para o primeiro teste obtém-se que:

Figura 14 - Estatísticas descritivas do ruído



Ou seja, os resíduos, aproximadamente, possuem média nula e uma distribuição muito próxima da distribuição normal.

Para o teste de variância constante, observa-se que:

Tabela 14 - Teste de White

Heteroskedasticity Test: White







































F-statistic	695.5274	Prob. F(3,245931)	0.0000	
Obs*R-squared	2069.061	Prob. Chi-Square(3)	0.0000	
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.569748	0.004408	129.2471	0.0000
MANHA	0.077864	0.006233	12.49264	0.0000
TARDE	0.177216	0.006232	28.43770	0.0000
NOITE	0.266713	0.006234	42.78490	0.0000
R-squared	0.008413	Mean dependent var	0.700210	
Adjusted R-squared	0.008401	S.D. dependent var	1.097313	
S.E. of regression	1.092694	Akaike info criterion	3.015185	
Sum squared resid	293636.5	Schwarz criterion	3.015354	
Log likelihood	-370765.8	Hannan-Quinn criter.	3.015234	
F-statistic	695.5274	Durbin-Watson stat	0.895430	
Prob(F-statistic)	0.000000			

O método utilizado foi o teste de White. Na literatura, o teste de White é construído de maneira muito intuitiva; ele consiste de uma regressão onde sua variável explicada é o resíduo ao quadrado do modelo original e os regressores são os mesmos do modelo original com adição de seus produtos cruzados e seus quadrados. Entretanto, como nosso modelo original só possui variáveis binárias, a inclusão dos termos cruzados e dos quadrados dos regressores deve ser evitada, uma vez que se pode recair sobre o erro de multicolinearidade entre os termos.

Sua interpretação é direta e segue o padrão da interpretação de um modelo de regressões qualquer. Pode-se perceber, portanto, que o modelo sofre com heterocedasticidade (a variância de sua série de ruídos não é constante) que é configurada por todas as variáveis que o constituem (vide coluna t-Statistic).

Por fim, testa-se a autocorrelação entre os resíduos:

Tabela 15 - Autocorrelação dos resíduos

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.761	0.761	142364	0.000
		2	0.483	-0.227	199806	0.000
		3	0.260	-0.046	216450	0.000
		4	0.107	-0.023	219262	0.000
		5	-0.006	-0.066	219272	0.000
		6	-0.112	-0.112	222369	0.000
		7	-0.199	-0.076	232072	0.000
		8	-0.271	-0.105	250106	0.000
		9	-0.319	-0.085	275126	0.000
		10	-0.355	-0.115	306132	0.000
		11	-0.385	-0.128	342529	0.000
		12	-0.427	-0.194	387375	0.000
		13	-0.375	0.062	421968	0.000
		14	-0.332	-0.196	449111	0.000
		15	-0.281	-0.088	468558	0.000
		16	-0.223	-0.095	480737	0.000
		17	-0.148	-0.069	486105	0.000
		18	-0.064	-0.093	487123	0.000
		19	0.016	-0.068	487182	0.000
		20	0.092	-0.071	489280	0.000
		21	0.183	0.019	497560	0.000
		22	0.311	0.114	521335	0.000
		23	0.462	0.193	573761	0.000
		24	0.577	0.130	655626	0.000
		25	0.453	-0.341	706032	0.000
		26	0.291	-0.009	726867	0.000
		27	0.155	-0.027	732796	0.000
		28	0.066	-0.018	733872	0.000
		29	0.000	-0.022	733872	0.000
		30	-0.060	0.006	734760	0.000

Pode-se perceber, portanto, que a série de resíduos possui um elevado grau de autocorrelação.

Dado que o modelo não foi bem sucedido em dois dos testes dirigidos à série de seus resíduos, é seguro afirmar que o mesmo não possui as propriedades derivadas do corolário do Teorema de Gauss Markov, expostos no início deste tópico. Entretanto, isto não significa que o mesmo é descartável, uma vez que estes testes (com exceção do último) não dizem muito sobre o poder de previsão do modelo em si.

O poder de previsão da série está intrinsicamente ligado com o R^2 do modelo estimado; quanto maior o seu valor, mais a variável explicada varia de acordo com os regressores do modelo. Como o R^2 do modelo escolhido é o maior dentre todos os modelos estudados, pode-se dizer que o mesmo é o melhor modelo ajustado à série. Entretanto, esta afirmativa poderia não ser verdadeira se outra variável explicativa fosse introduzida ao modelo.

4.3.5. Intervalos de confiança e classificação

O uso de MQO é apenas uma parte do algoritmo, a outra se trata da classificação em si, que utiliza os consumos previstos pelo modelo juntamente com um intervalo de confiança para a obtenção dos potenciais clientes.

Tendo em mente que o projeto se interessa apenas pelos clientes que apresentam picos anormais durante o verão, o intervalo de confiança pode ser construído da seguinte maneira:

$$\hat{p}_i + \text{percentil } 80\%(u)$$

Onde \hat{p}_i representa o valor predito de p_i ; u a distribuição de resíduos produzidos pelo modelo. Escolhe-se trabalhar com percentis ao invés da variável normal, uma vez que a série de resíduos não apresenta a mesma distribuição.

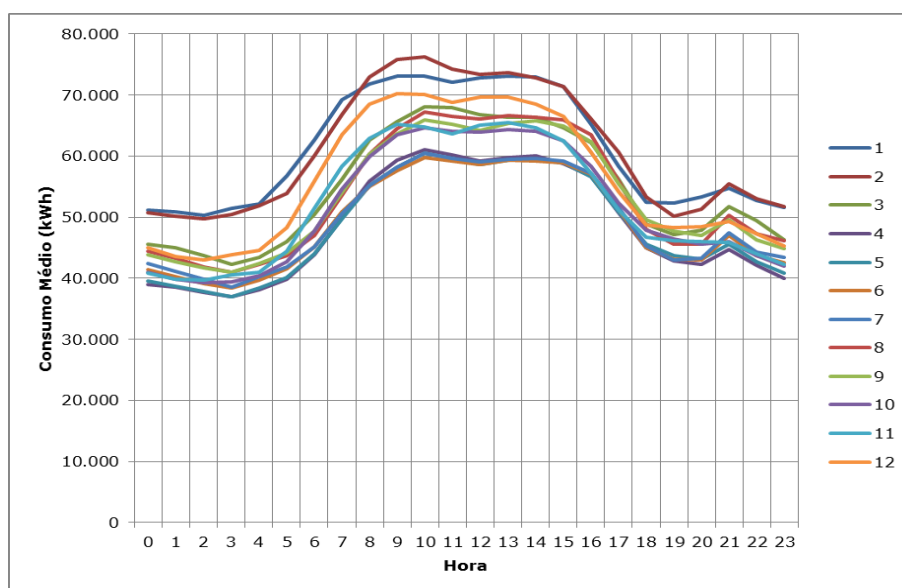
Desta forma, assume-se que qualquer cliente com classificações acima do intervalo de confiança definido, para qualquer um dos períodos da manhã e da tarde durante o verão ou primavera, possuem o perfil desejável para adoção de sistemas de energia fotovoltaicas em sua rede.

Para executar esta heurística, uma planilha Excel foi criada com todos os dados necessários. Uma descrição detalhada é feita de cada coluna:

- Aba 'Series':
 - NORM_TRUE: Dados de consumo normalizados, p_i .
 - PREDICTED: \hat{p}_i estimado pelo modelo.
 - PREDICTED_LIMIT: Coluna anterior com a adição dos limites para cada observação.
 - Flag: Marca um (1) se valor real for maior do que o registrado na coluna anterior; zero (0) caso contrário.
 - Serial: Número de identificação do cliente.
 - Manhã: Variável binária que demarca o período da manhã.
 - Tarde: Variável binária que demarca o período da tarde.
 - Noite: Variável binária que demarca o período da noite.
 - Verão: Variável binária que demarca os meses de verão.
 - Outono: Variável binária que demarca os meses de outono.
 - Verão: Variável binária que demarca os meses de verão.
 - Primavera: Variável binária que demarca os meses de primavera.
 - Hora: Variável que contém informação da média horária do cliente, h_i .
- Aba 'Lim':
 - Residual: Coluna que guarda a informação dos resíduos gerados pelo modelo.
 - Limite: Armazena o valor de octogésimo percentil da série dos resíduos.
- Aba 'TABELA_DINAMICA':
 - Contém uma tabela dinâmica onde pode-se verificar a quantidade de violações do limite para cada cliente; no período da tarde ou da manhã; nas estações de verão ou primavera.
- Aba 'Output':
 - Serial: Número de identificação do cliente.
 - Primavera: Marcador de pontos na primavera.
 - Verão: Marcador de pontos no verão.
 - Cliente Fotovoltaica: Marca um (1) se o cliente tem perfil favorável ou zero (0) caso contrário.

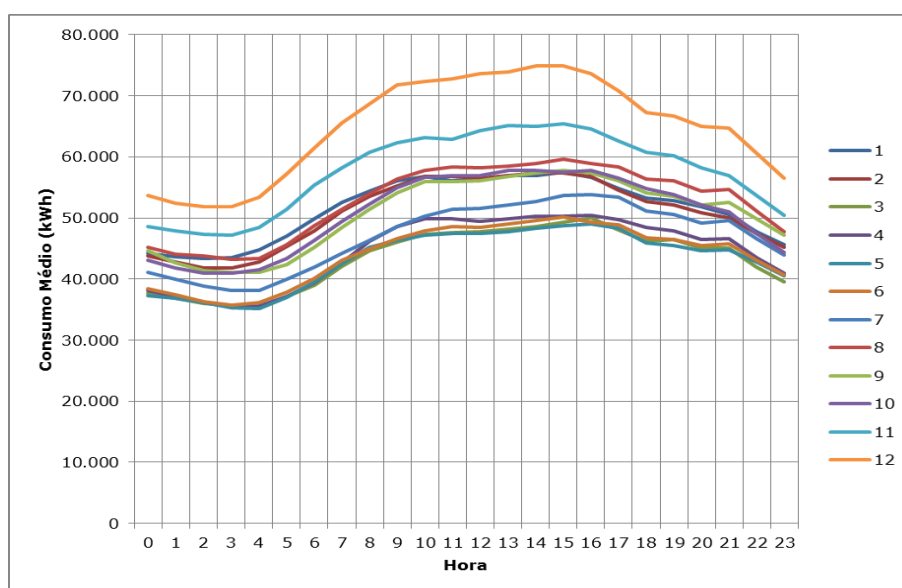
Finalmente, a quantidade de potenciais clientes classificados, dentre os 2.212 contidos na amostra, é 1.318. A média horária agregada, para cada mês do ano, é exibida abaixo:

Figura 15 - Médias horárias dos clientes 'potenciais'; Meses na em ordem numérica



Comparativamente, pode-se exibir o mesmo gráfico para os clientes 'não potenciais':

Figura 16 - Médias horárias dos clientes 'não potenciais'; Meses na em ordem numérica



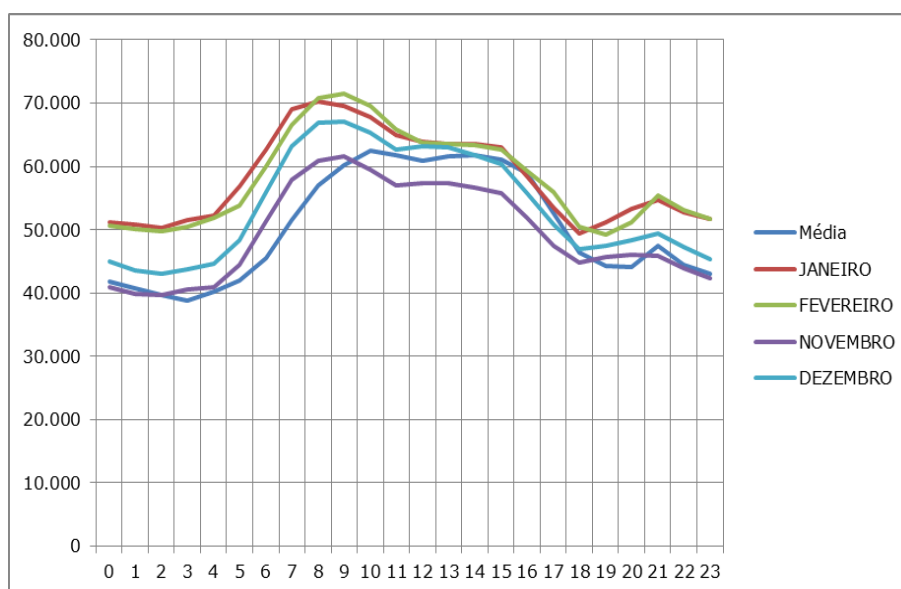
De imediato, percebe-se que o gráfico dos clientes 'não potenciais' apresenta uma variação de consumo menor durante o dia, o que gera uma curva com inclinações mais suaves se comparada à curva dos clientes 'potenciais'. Isto se deve ao fato de que o algoritmo explora as diferenças entre o consumo médio do cliente e seu consumo na ponta: quando esta for suficientemente grande, o mesmo pode ser um potencial contratante do serviço. Quanto ao nível de consumo, percebe-se que é bastante parecido entre 'potenciais' e 'não potenciais'.

5. Análise de investimento

5.1. Análise de Custos e Eficácia

Como apontado no capítulo introdutório desta monografia, o objetivo final do estudo é a redução das curvas agregadas de consumo, associadas aos meses de novembro, dezembro, janeiro e fevereiro, dos clientes classificados, ao mesmo patamar da média dos meses complementares. Desta forma, obtém-se:

Figura 17 - Análise de eficácia



Este nível de redução equivale à adição de 60.000 painéis fotovoltaicos com as especificações descritas no capítulo "Tratamento de Dados", tópico "Painel Escolhido". Portanto, como estabelecido no capítulo introdutório desta monografia, o objetivo pode ser atingido apenas com a aplicação de painéis fotovoltaicos às redes de energia dos consumidores classificados da concessionária.

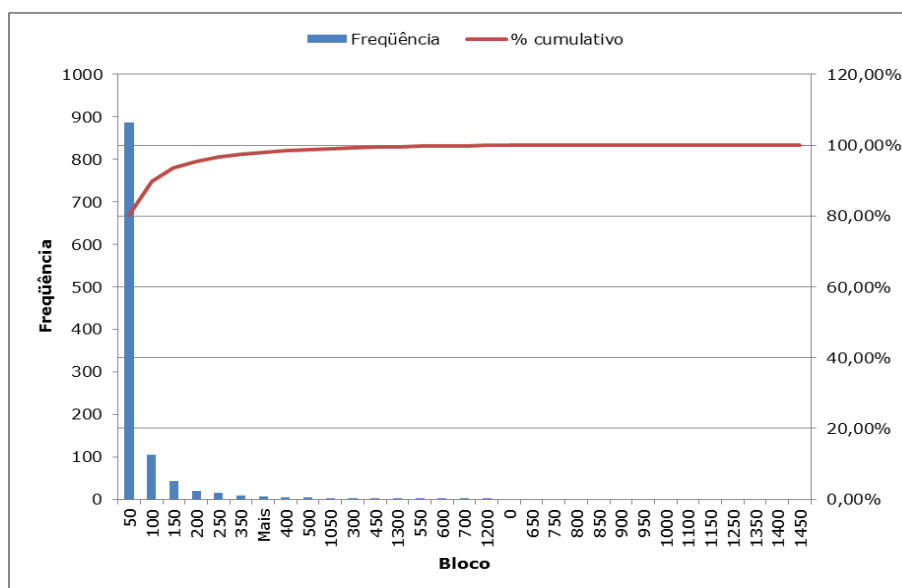
Ademais, pode-se afirmar que, sem contar com gastos de mão de obra e obra civil, o custo deste projeto equivaleria a R\$ 132.000.000,00 o que, em média, representa um ônus de R\$ 100.151,75 por cliente. Naturalmente, esta média não pode ser aplicada a todos, uma vez que cada um possui um nível de consumo diferente.

5.2. Análise de Pareto

Assumindo que todos os clientes escolhidos devem contribuir com a mesma proporção energética, i.e., economizar o mesmo percentual de energia, pode-se definir a quantia exata do investimento necessário por cada um. Utilizando-se da máxima de cada cliente nos meses de interesse, o objetivo é reduzir seu consumo para a média dos meses complementares, como no caso agregado.

Desta forma, obtém-se a seguinte distribuição para estes investimentos, em quantidade de painéis comprados (histograma na forma de Pareto):

Figura 18 - Histograma de Pareto



É evidente que a maior parte dos clientes (cerca de 80%), necessita de um investimento menor do que 50 painéis, o que significa que apenas 20% dos clientes deverão assumir custos, aproximadamente, maiores do que a média calculada de R\$ 100.151,75.

Portanto, executando-se uma análise de Pareto (regra 80/20) dos dados obtidos:

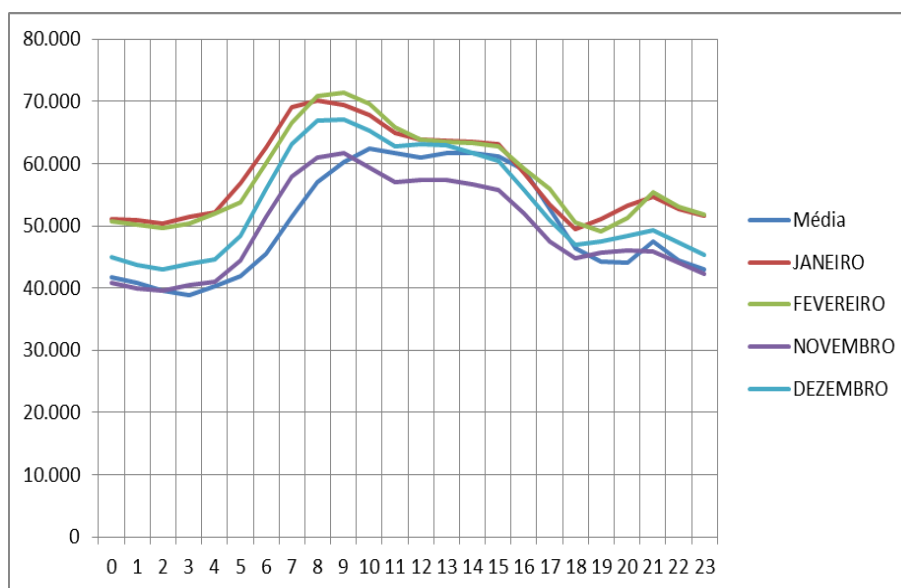
Tabela 16 - Análise de Pareto

	20% da amostra	80% da amostra
Número de painéis agregados	50.731	12.815
Energia acumulada gerada em pico kW/dia	55	3
C. Máx kW/dia (Jan, Fev, Nov, Dez)	256	22
C. Médio kW/dia (Meses Complementares)	201	18
Custo	R\$ 111.608.200,00	R\$ 28.193.000,00
Custo Médio	R\$ 511.964,22	R\$ 25.490,96

Vê-se que a regra 80/20 é aplicável ao problema, mostrando que certa de 5/6 de todo o investimento deve ser focado em grandes clientes que consomem, em média, 256kW de energia por dia. Portanto, investindo em apenas 20% dos clientes classificados pelo algoritmo, será possível alcançar 79,8% da economia desejada. Ademais, pode-se afirmar que, dos 1.318 clientes classificados, o investimento seria particularmente eficaz se focalizado em apenas 218 destes.

Com este nível de investimento, exibe-se, novamente, a curva de economia desejada:

Figura 19 - Análise de eficácia como nova quantidade de painéis



Como previsto, apesar de menos painéis, o investimento ainda surtirá efeito considerável na redução de risco do sistema, uma vez que foca nos clientes que mais consomem energia.

5.3. Retorno de investimento

O próximo passo é verificar o tempo de retorno do investimento para estes clientes, a fim de que seja auferida a atratividade do mesmo. Para tanto, necessita-se do custo médio de energia por kWh e da energia gerada pelo painel, juntamente com seu preço. Utilizando-se as tabelas exibidas no capítulo "Tarifas Vigentes" pode-se deduzir que a taxa R\$/kWh da Light é, em média, R\$234,15/MWh, ou R\$0,23/kWh. Acrescendo-se o custo de sua carga tributária – PIS, COFINS e ICMS – têm-se: R\$0,38/kWh, em média. Por ano, utilizando-se os dados de insolação solar obtido do INMET, estima-se que um painel solar do tipo exibido nesta monografia é capaz de gerar cerca de 380 kWh de energia por ano. Isto equivale a uma economia de R\$144,00 reais anuais. Sem considerar a projeção anual da inflação (moeda constante), cada painel solar se paga, em média, após 15 anos de uso.

Considerando os custos financeiros, ao aplicarmos o valor de um painel solar na poupança, com juros reais de, em média, 0,6% ao ano⁷ [FONTE UOL NOTÍCIAS], este período passa a ser 17 anos.

6. Conclusões

6.1. Conclusões

Através do estudo verificou-se que é possível estruturar um plano de cogeração de energia fotovoltaica, agindo em poucos clientes chave, para diminuir o risco de sobrecargas no sistema da Light. No entanto, este empreendimento se mostra inviável financeiramente, uma vez que seu retorno é extremamente fraco, alcançando os juros reais ínfimos da poupança em 17 anos.

6.2. Comentários Finais

Para melhorar este aspecto, sugere-se a utilização de um painel fotovoltaico com um rendimento maior do que o utilizado neste estudo, alcançando um custo benefício melhor (menor preço, maior geração de energia). Ademais, como visto em países do exterior, como Portugal e EUA, o governo poderia incentivar este tipo de investimento através de incentivos fiscais ou de crédito especiais. Nestes países, por exemplo, o retorno de investimento seria alcançado em cerca de 11 anos através de um incentivo fiscal, praticado em ambos, que deduz em 30% o preço do equipamento.

Sobre o aspecto algorítmico e de classificação, um passo seguinte seria eliminar os efeitos indesejados sobre o ruído da regressão, melhorando seu poder de previsão e ajustando melhor as classificações de potenciais clientes. Como neste estudo não foram considerados fins de semana e feriados para o cálculo da média, este seria um ponto imediato de melhoria. Ademais, pode-se destacar a influência que outras variáveis podem exercer sobre poder preditivo da regressão, como temperatura ambiente, índice IPCA, crescimento do PIB estadual dentre outras.

7. Bibliografia

[1] Disponível em: < :http://www.aneel.gov.br/area.cfm?id_area=65 >. Acesso em: 29 jun. 2013.

[2] Disponível em: < :http://ri.light.com.br/ptb/925/Press_Release_1T11_Final.pdf >. Acesso em: 29 jun. 2013.

Disponível em: < :http://ri.light.com.br/ptb/2522/Press_Release_1T12_SA_Final.pdf >. Acesso em: 29 jun. 2013.

Disponível em: < :http://ri.light.com.br/ptb/3462/Press_Release_1T13_Final.pdf >. Acesso em: 29 jun. 2013.

[3] Queiroz, P.V.S, *Mensuração do consumo de energia elétrica: algoritmo para detecção de potenciais usuários de termoacumulação como alternativa para deslocamento de carga*. Dissertação de Mestrado em Metrologia, Pontifícia Universidade Católica do Rio de Janeiro, Defesa em 19 de setembro de 2011.

[4] Disponível em: < :<http://www.light.com.br/para-residencias/Sua%20Conta/composicao-da-tarifa.aspx> >. Acesso em: 29 jun. 2013.

[5] Disponível em: < :<http://www.cresesb.cepel.br/sundata/> >. Acesso em: 29 jun. 2013.

[6] Disponível em: < :<http://www.kyocerasolar.com.br/site/arquivos/produtos/76.pdf> >. Acesso em: 29 jun. 2013.

[7] Disponível em: < :<http://economia.uol.com.br/noticias/redacao/2013/01/10/rendimento-da-poupanca-em-2012-e-o-menor-em-46-anos-diz-consultoria.htm> >. Acesso em: 29 jun. 2013.