

Detecção de *outliers* no consumo de energia elétrica

Ivana Soares Guarany, Raimundo Carlos Silvério Freire, Andressa da Silva e Edmar Candeia Gurjão

Resumo—A identificação de comportamentos anômalos no consumo de energia pode indicar defeitos, consumo não previsto, e oportunidades para uso de fontes alternativas e armazenadores de energia. A análise do consumo é o primeiro passo para se identificar ações que resultem em eficiência energética e redução do consumo de energia. Neste trabalho se propõe uma abordagem para identificar consumo anômalo de energia elétrica utilizando uma abordagem estatística baseada em interquartís para a definição de *outliers*. O método foi testado utilizando os dados de consumo de energia fornecidos por um medidor inteligente instalado em um prédio da Universidade Federal de Campina Grande (UFCG). Os resultados mostram que é possível a utilização do método para a detecção automática de anomalias.

Palavras-Chave—*Outlier*, Consumo, Gerenciamento de Energia, Detecção de Anomalias.

Abstract—Identification of outliers in electrical energy consumption may indicate defects, non authorized consumption, and opportunities for utilization of alternative energy generators and energy accumulators. The consumption analysis is the first step to identify these outliers and actions that results in energy efficiency and reduction in consumption. In this paper it is proposed a statistical method to identify the outliers. Using data from an smart meter installed in a building of the Federal University of Campina Grande the proposed method detect the outliers with a high accuracy.

Keywords—*Outlier*, Consumption, Power Management, Detection of Anomalies.

I. INTRODUÇÃO

Analisar os dados de consumo de energia elétrica permite a detecção de defeitos, entendimento do padrão de consumo e oportunidades para melhoria da eficiência energética [8]. Neste contexto, a identificação de consumos que diferem dos normalmente observados é uma oportunidade para detectar os que não são autorizados, provocando diminuição de custos [11].

Segundo [3], anomalia, doravante denominada *Outlier* por estar consagrado na literatura da área, pode ser definida como um valor que desvia tanto dos outros observados de tal forma que possa ter sido gerado por causas diferentes. Diversas técnicas tem sido desenvolvidas para a detecção de anomalias. No caso da energia elétrica, essas técnicas podem ser utilizadas para identificar quando o consumo desvia muito do que pode ser considerado o padrão da curva de consumo de energia elétrica e, assim, indicar um comportamento anômalo, o que pode ajudar na redução de custos ao indicar o momento temporal de ocorrência e as possíveis causas (*e.g.* mau funcionamento de equipamentos, queda de fases, iluminação em horários inadequados, atividades fora do horário normal) [6].

Ivana Soares Guarany, Raimundo Carlos Silvério Freire, Andressa da Silva e Edmar Candeia Gurjão, Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande, Campina Grande-PB, E-mails: ivana.guarany@ee.ufcg.edu.br, rcfreire@dee.ufcg.edu.br, andressa.silva@ee.ufcg.edu.br e ecg@dee.ufcg.edu.br.

Neste artigo é apresentado um método de detecção automática do comportamento anômalo do consumo de energia com o intuito de otimizar o consumo e reduzir a necessidade da verificação visual de todo o conjunto de dados. Como exemplo da aplicação, é utilizado o consumo em um dos prédios da Universidade Federal de Campina Grande (UFCG) - Campus Campina Grande. O objetivo principal do trabalho é apresentar uma abordagem baseada em detecção de *outliers* para identificar quando o consumo de energia de um prédio é diferente do que pode ser considerado o comportamento padrão, indicando se o consumo medido no tempo atual é normal ou está acima ou abaixo do intervalo de valores esperados.

II. GERENCIAMENTO DE ENERGIA

O aumento na demanda de energia elétrica está relacionada a sua utilização nos mais diversos setores, o que traz benefícios econômicos e melhoria da qualidade de vida, por outro lado também implica em esgotamento dos recursos utilizados para a produção de energia, quando utilizadas fonte não renováveis, e em impacto ao meio ambiente [1][5][11]. Tem-se estimulado o uso eficiente da energia elétrica, porém o modelo tradicional do sistema elétrico e do registro de informações de consumo, não propicia aos consumidores informações suficientes que permitam decisões mais efetivas a respeito desse consumo.

Uma das opções é a utilização de dispositivos e sistemas de gerenciamento energético que forneçam informações detalhadas sobre o consumo de energia. Ao conhecer o consumo, fica mais fácil determinar opções que possam torná-lo eficiente, por exemplo, detectando e eliminando comportamentos anômalos [13].

O gerenciamento do consumo de energia elétrica normalmente utiliza medidores inteligentes, fontes alternativas de energia, armazenadores e uma plataforma que consiga monitorar, controlar e atuar nas transferências de carga entre a rede elétrica da concessionária e a energia disponível localmente. Nas próximas seções são apresentadas as técnicas utilizadas neste trabalho para a detecção de anomalias.

III. IDENTIFICAÇÃO DE OUTLIERS

Segundo [3], *Outlier* pode ser definido como um valor que desvia tanto dos outros observados que pode ter sido gerado por causas diferentes. De acordo com [10], *Outliers* podem ser classificados como:

- pontuais: são valores ou conjuntos de valores que diferem de todos os outros pertencentes ao conjunto de dados, inclusive entre si. Um exemplo é um valor de consumo muito elevado ou muito baixo que ocorreu poucas vezes ou uma vez.

- contextuais: são chamados também de *outliers* condicionais, pois são instâncias que podem ser anômalas ou não, dependendo do contexto. Um exemplo é um valor de consumo baixo em horário de pico. Apesar do valor ser considerado normal em outro horário, no contexto do horário de pico o valor é anormal. O mesmo vale para o caso de um valor elevado for observado em um horário que não é de pico.
- coletivos: são conjunto de valores que apresentam as mesmas características entre si, mas diferem de todo o resto do conjunto de dados. Desta forma, surge a consequência de que a correlação entre as instâncias é necessária para que possa existir os *outliers* coletivos, ao contrário dos *outliers* pontuais. Como exemplo desse tipo de anomalia, está o caso onde um conjunto de valores assume características inesperadas por um longo tempo.

A identificação de *outliers* está sendo bastante estudada e aplicada para diversos fins [12]. Existem diversos métodos de identificação de *outliers* e a escolha pode depender de vários fatores, como o tipo do dado e o tamanho do conjunto de dados, além da interpretabilidade (*interpretability*) do *outlier*, que se refere às possíveis interpretações, fornecendo indícios e descrições das causas [2] [9]. Desta forma, no contexto deste artigo, o método escolhido e o conhecimento produzido podem ajudar no gerenciamento da energia elétrica.

Alguns métodos de identificação que podem ser citados são os baseados em classificação, *clustering*, *nearest neighbors*, Teoria da Informação, análise espectral, baseados em densidade e métodos estatísticos [9] [10].

Neste trabalho, é adotada uma abordagem estatística baseada em quartis. Quartil é o nome que se dá a um dos 3 valores que dividem um conjunto ordenado de valores em 4(quatro) partes iguais. Estando os dados de forma ordenada, encontra-se a mediana e os valores máximo e mínimo da amostra. A mediana dos valores entre o menor valor e a mediana de todos os valores é chamada de primeiro quartil ou quartil inferior. A mediana dos valores entre a mediana de todos os valores e o valor máximo é chamada de terceiro quartil ou quartil superior. Utiliza-se os quartis superior Q_3 e inferior Q_1 e a diferença entre eles, $IQR = Q_3 - Q_1$, que é chamada de intervalo de interquartil. Todos os valores que estiverem acima de $Q_3 + 1,5(IQR)$ ou abaixo de $Q_1 - 1,5(IQR)$ são definidos como *outliers* [4]. A região entre esses intervalos contém 99,3% das observações. Existem abordagens usando esse princípio em campos como anomalias em dados médicos e anomalias em turbinas de motores [10]. Esta abordagem traz as características interessantes como simplicidade de implementação e a capacidade de identificar múltiplos *outliers* [7].

De posse de um vetor de amostras numéricas, os passos utilizados para identificar os *outliers* são os seguintes:

Passo 1: Ordenar os números em ordem crescentes

Passo 2: Identificar a mediana das amostras

Passo 3: Identificar o maior e o menor valor

Passo 4: Identificar a mediana entre o menor número e a mediana geral de toda a série de dados (Q_1), e a mediana entre a mediana geral e o maior número na série (Q_3)

Passo 5: Calcular o intervalo de interquartil ($IQR = Q_3 - Q_1$)

Para o caso analisado neste trabalho, feitas n medições repetidas m vezes, tem-se a matriz de dados

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$$

em que cada elemento x_{ij} representa a j -ésima medição da i -ésima repetição.

Seja um vetor $X_{.j}$, um vetor coluna $m \times 1$ de \mathbf{X} , ele denota os valores de consumo medidos no j -ésimo instante em todas as m semanas.

$$\mathbf{X}_{.j} = [x_{1j} \quad x_{2j} \quad \dots \quad x_{mj}]^T$$

Os vetores coluna são, então, processados pela função de cálculo de *outliers* que, por sua vez, retorna as k vezes que foram classificados como *outliers*.

IV. METODOLOGIA

Para a aplicação do método, foram utilizados dados fornecidos por um medidor trifásico desenvolvido na UFCG - Campus Campina Grande pelo Grupo de Pesquisa em Processamento de Sinais e Informação. O medidor fica conectado à alimentação do prédio e mede as potências ativa e reativa, além do fator de potência. Os valores medidos consistem em uma média de 60 medições por segundo. A cada minuto, essa média é enviada via Internet para um servidor onde as informações são armazenadas em um banco de dados, e que podem ser acessadas via *web*. A taxa de 1 minuto escolhida se justifica para análise mais detalhada dos eventos, como variações de consumo e interferências momentâneas causadas por equipamentos com defeito. O medidor desenvolvido está em fase de validação e ainda não possui certificação de calibração pelo INMETRO.

De posse dos dados, é realizada a análise para classificar os períodos de consumo. O critério adotado para classificar um conjunto de valores de consumo como anômalos foi esse conjunto ser composto por pelo menos 5 (cinco) *outliers* consecutivos. Desta forma, o tempo especificado para rotular um conjunto de valores como anômalos foi de 5 minutos. A Fig. 1 apresenta o fluxograma do método de identificação de anomalias no consumo de energia.

Há duas possibilidades de utilização da metodologia empregada. A primeira é uma determinação de *outliers* de forma *offline*, em que não é obrigatório o passo de classificação dos sinais, que fica ao critério do usuário. Neste caso, o algoritmo informa os *outliers* globais sem requerer que os sinais apresentem um comportamento padrão. A segunda possibilidade é a determinação em tempo real, quando o passo de classificação é crucial, pois é necessário um conjunto de dados bem definidos para fazer os cálculos dos limites inferior e superior de forma a representar uma curva padrão. Nos dois casos, se 5 (cinco) medidas consecutivas forem *outliers*, o conjunto é classificado como um comportamento anômalo e um alarme é gerado.

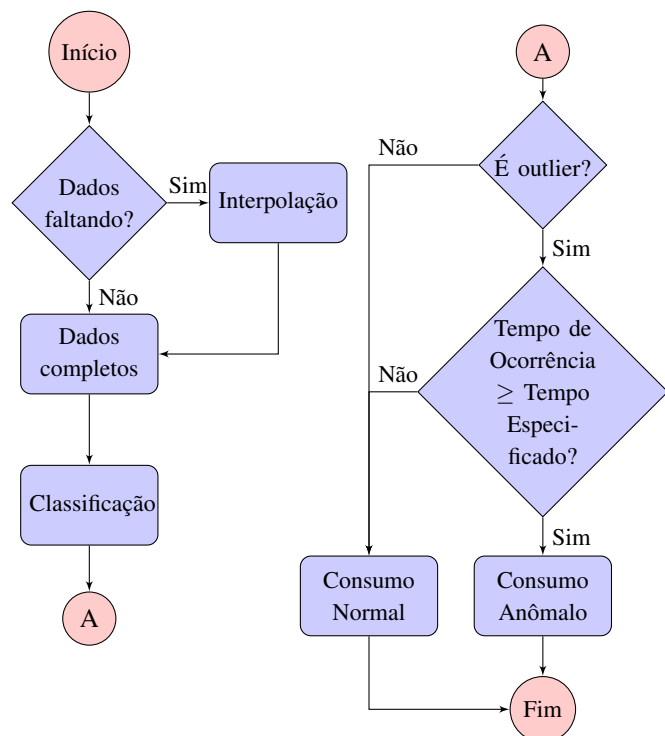


Fig. 1. Fluxograma da detecção de anomalias.

V. RESULTADOS E DISCUSSÃO

Como estudo de caso, foram utilizados os dados do prédio da UFCG denominado Bloco CJ (coordenadas geográficas -7.212135, -35.908530), que é um ambiente com diversas salas e que são utilizadas para as mais variadas atividades como aulas, laboratórios de pesquisa e salas de professores. Os primeiros passos foram a coleta dos dados referentes às medições do consumo e o tratamento desses dados. Estando os dados tratados, foi feita uma classificação visual dos padrões de consumo e, a partir da classificação, foram feitos os cálculos associados à definição dos *outliers*. O conjunto de dados foi separado por dias da semana para a identificação, pois de um dia para o outro não foi observado um padrão definido.

Os principais equipamentos instalados no Bloco CJ são ar-condicionados, lâmpadas e computadores. A existência ou não de padrões está associada ao uso do ar-condicionado nos horários nos dias em que há aula no prédio, a presença dos professores em suas salas ou a presença de pessoas nos laboratórios. A Fig. 2 apresenta as curvas de consumo do Bloco CJ para os dias de quarta-feira para cada uma das 7 semanas estudadas.

A partir da Fig. 2, é possível observar a presença de dois intervalos de tempo em que há maior consumo de energia, e que são comuns à maioria das curvas. Esses intervalos, definidos por (600, 650) (em minutos) e (950, 1060) (em minutos), correspondem aproximadamente aos intervalos entre 10h e 10h50 e 15h50 e 17h40 respectivamente. Esses horários estão relacionados aos períodos em que há presença de pessoas trabalhando ou aulas sendo realizadas. A presença do maior consumo nestes intervalos se repete na maioria dos dias de segunda a sexta, mas com características como média,

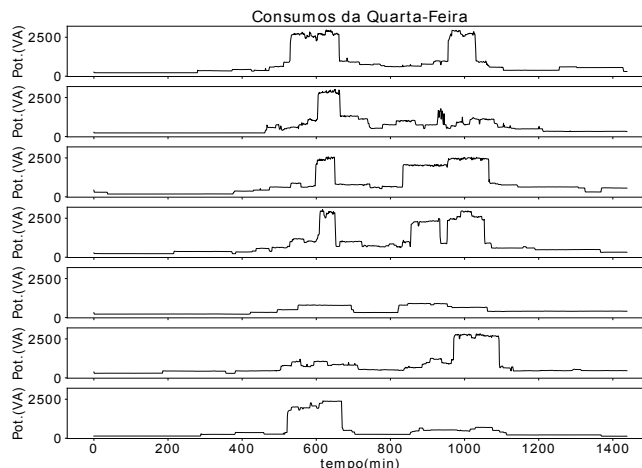


Fig. 2. Curvas de consumo no Bloco CJ para os dias de Quarta das 7 (sete) semanas estudadas.

variância e duração específicas para cada dia. Também é possível observar que em alguns dias há um consumo elevado nos intervalos (530, 650) (em minutos) e (833, 1060) (em minutos), que correspondem aproximadamente aos horários de 8h40 a 10h50 e 13h50 a 17h40. Isso se dá pelo mesmo motivo explicitado anteriormente.

A técnica de detecção de anomalias descrita foi aplicada nos dados de 7 semanas de medições feitas no Bloco CJ. A Fig. 3 representa as curvas de consumo do Bloco CJ para os dias de quarta-feira de 7 (sete) semanas distintas e os intervalos de consumos anômalos identificados automaticamente. Os retângulos indicam os intervalos anômalos detectados e as linhas tracejadas correspondem aos consumos do dia para todas as semanas. Os retângulos vermelhos indicam os intervalos em que os valores de consumo se encontram acima do limite superior determinado por $Q_3 + 1,5(IQR)$ e com duração maior ou igual ao tempo mínimo especificado. Os retângulos azuis indicam os intervalos em que os valores de consumo se encontram abaixo do limite inferior determinado por $Q_1 - 1,5(IQR)$ e com duração maior ou igual ao tempo mínimo especificado.

Apesar de, aparentemente, um dia ter um consumo bem menor que os outros, o algoritmo identificou somente alguns intervalos como anômalos. Isso pode ser explicado recorrendo ao modo de como os intervalos usados para a identificação são definidos. Quanto maior o valor do intervalo de interquartil (IQR), maior será o valor do intervalo que define os valores tidos como padrões.

Desta forma, nos intervalos (700, 850) (em minutos) e (1100, 1440) (em minutos), onde os valores de um dia estão bem abaixo dos demais, espera-se valores de IQR altos, pois os sinais nesses intervalos apresentam valores que se distanciam um dos outros. Assim, apesar do consumo de um dia estar visualmente abaixo dos demais, ainda assim os valores de consumo desta dia são classificados como normais. O contrário ocorre no intervalo (0, 220) (em minutos), pois visualmente alguns valores parecem normais. Entretanto, como os valores apresentam médias próximas uma das outras, o

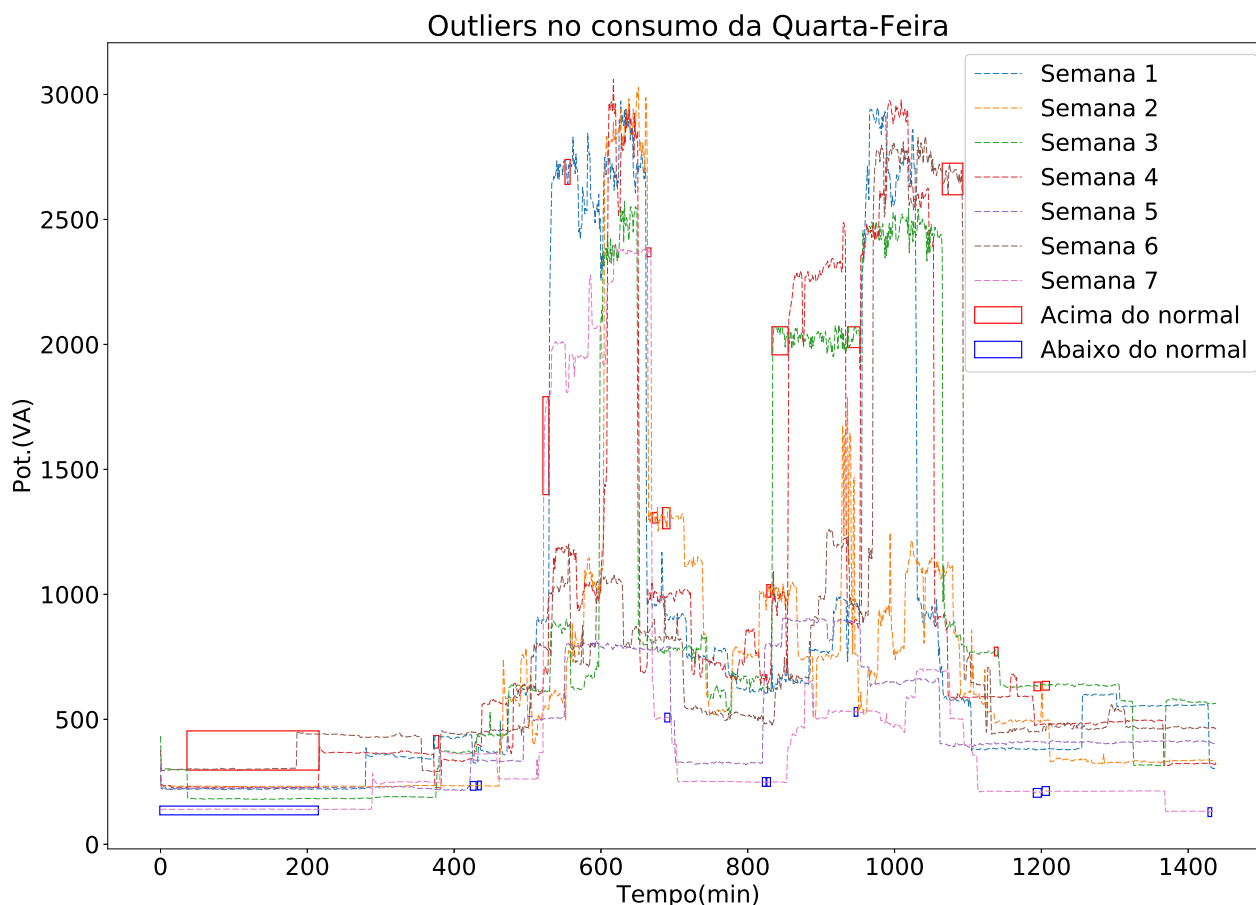


Fig. 3. Curvas de consumo no Bloco CJ para o dia de quarta-feira (tracejado) de 7 semanas e intervalos de consumos anômalos (retângulos).

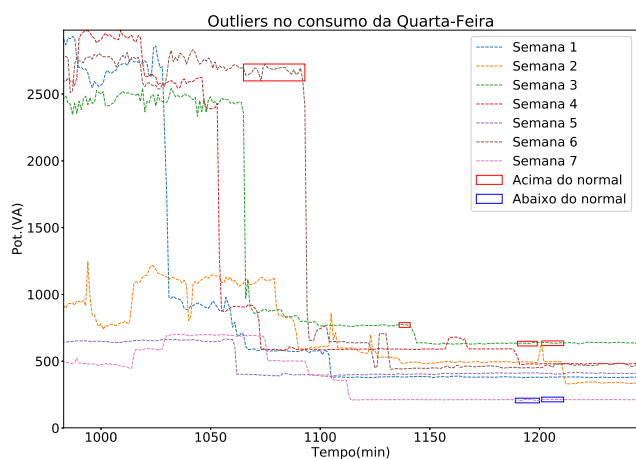


Fig. 4. Curvas de consumo no Bloco CJ para o dia de quarta-feira (tracejado) de 7 semanas e intervalos de consumos anômalos (retângulos).

valor de IQR tende a ser menor e, portanto, os intervalos que definem os *outliers* também diminuem, fazendo com que haja a demarcação de dois intervalos como anômalos, um acima do consumo normal e outro abaixo. Os *outliers* identificados

podem ser interpretados como contextuais, pois é apresentado um consumo possível, mas com um valor não comum para o horário.

A identificação de *outliers* contextuais pode ser muito importante para o usuário por indicar um consumo que foge dos padrões para determinados horários. Em consumos abaixo do normal, pode indicar uma subtensão ou diminuição na atividade no horário. Para consumos acima do normal, há um acréscimo nos gastos e, portanto, uma indicação automática deste comportamento pode ajudar na investigação das causas do consumo elevado, o que ajuda na redução de custos. A indicação pode também ajudar a identificação de defeitos em equipamentos, sobretensão ou utilização do prédio em horários incomuns. O segundo caso pode ter uma aplicação mais imediata, como o desligamento de ar-condicionados ou outros equipamentos que, porventura, foram esquecidos ligados. Na Fig. 3, podem ser observadas algumas identificações do aumento no consumo em um tempo anterior ao comum (acionamento precoce) e a permanência do consumo elevado por um tempo maior do que o comum (desligamento tardio). Na Fig. 4 é apresentado, mais detalhadamente, um intervalo de identificação de um desligamento tardio.

No caso dos finais de semana, os *outliers* identificados

foram os coletivos, pois o contexto, neste caso, não estava bem definido. Isso se deve ao fato de que nos finais de semana não foram observados padrões de consumos característicos. As Fig. 5 e Fig. 6 apresentam as curvas de consumo dos sábados e domingos das 7 semanas estudadas.

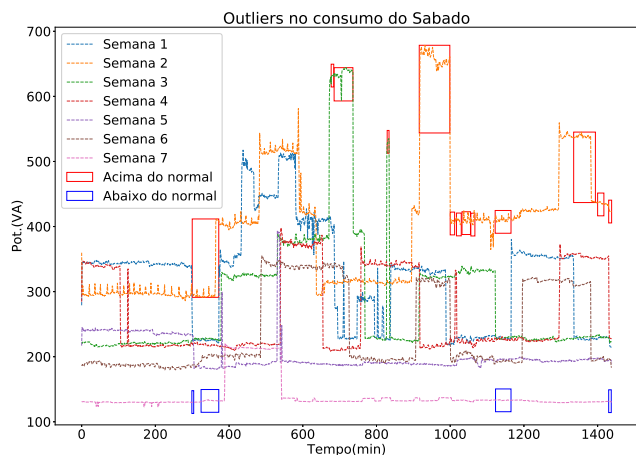


Fig. 5. Curvas de consumo no Bloco CJ para os sábados (tracejado) de 7 semanas e intervalos de consumos anômalos (retângulos).

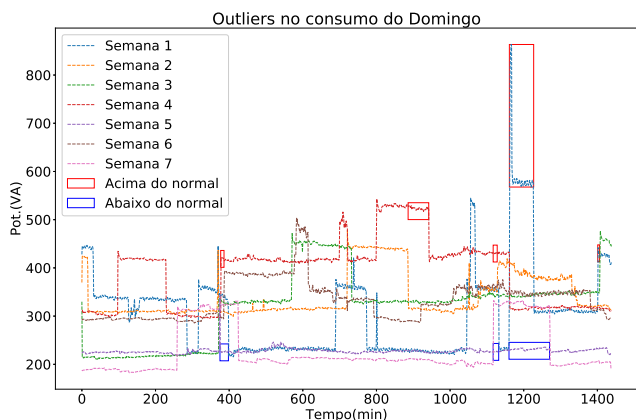


Fig. 6. Curvas de consumo no Bloco CJ para os domingos (tracejado) de 7 semanas e intervalos de consumos anômalos (retângulos).

Alguns intervalos dos consumos estão visualmente abaixo ou acima da maioria, mas somente alguns são classificados como anômalos. Isso ocorre pelo que já foi explicado acerca da definição do valor de IQR. Alguns intervalos de consumo estão claramente acima dos demais e foram identificados pelo algoritmo como anômalos. Esses consumos acima do normal indicam algum tipo de atividade incomum, pois, geralmente nos finais de semana, as atividades no prédio são reduzidas.

Como pode ser observado, as curvas de consumo do prédio não apresentam um padrão bem definido. Uma abordagem simples capaz de identificar comportamentos anômalos mesmo nestas condições se torna uma solução interessante. É possível utilizar os resultados da classificação para determinar se um dia inteiro difere dos demais a partir do estabelecimento de um limiar de quantas anomalias por dia foram observadas.

VI. CONCLUSÃO

A detecção de anomalias é uma das ações para o gerenciamento eficaz da demanda de energia em uma unidade consumidora. Desta forma, este trabalho propôs uma técnica para detecção de anomalias no consumo de energia elétrica. O método consistiu na utilização da detecção de *outliers* por meio da abordagem estatística de interquartis e foi utilizada como fonte de dados o consumo de energia em um prédio da UFCG.

A UFCG é um perfil consumidor diferenciado, pois possui consumo em horários diversos, sem uma periodicidade fixa que permita definir horários invariantes de funcionamento e picos de demanda, haja vista que a instituição possui atividades diversas em horários alternados, e os semestres letivos nem sempre se encaixam nos mesmos meses do ano. Portanto, identificar padrões de consumo nesses tipos de consumidores é uma tarefa complexa e, conseqüentemente, definir o que foge do padrão também o é.

Este trabalho apresentou um método automático, empiricamente rápido e que utiliza uma abordagem estatística robusta para a identificação de anomalias de consumo de energia em um prédio que não apresenta uma curva de consumo tão bem delineada. Como trabalho futuro, pretende-se comparar o método aplicado aqui com outros a fim de determinar a eficiência. Pretende-se também explorar a identificação de comportamentos anômalos em tempo real.

AGRADECIMENTOS

Os autores agradecem o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERÊNCIAS

- [1] B. Chen, M. Sinn, J. Ploennings e A. Schumann. "Statistical Anomaly Detection in Mean and Variation of Energy Consumption", *International Conference on Pattern Recognition*, pp. 3570–3575, 2014.
- [2] C.C. Aggarwal. *Outlier Analysis*, 2.ed. Springer, 2017.
- [3] D. Hawkins. *Identification of Outliers*, Chapman and Hall, London, 1980.
- [4] D.C. Hoaglin, B. Iglewicz e J.W. Tukey. "Performance of Some Resistant Rules for Outlier Labeling", *Journal of the American Statistical Association*, v. 81, n. 396, pp. 991–999, 1986.
- [5] H. Chen, X. Fei, S. Wang, X. Lu, G. Jin e X. Wu. "Energy consumption Data Based Machine Anomaly Detection", *Second International Conference on Advanced Cloud and Big Data*, pp. 136–142, 2014.
- [6] Jui-Sheng Chou e A.S. Telaga. "Real-time detection of anomalous power consumption", *Renewable and Sustainable Energy Reviews*, n. 33, pp. 400–411, 2014.
- [7] J.W. Tukey. *Exploratory Data Analysis*, Addison Wesley, 1977.
- [8] K. Vikhorev, R. Greenough e N. Brown. "An advanced energy management framework to promote energy awareness", *Journal of Cleaner Production*, vol. 43, pp. 103–112, 2013.
- [9] M. Gupta, J. Gao, C. Aggarwal e J. Han. *Outlier Detection for Temporal Data*, Morgan & Claypool, 2014.
- [10] V. Chandola, A. Banerjee e V. Kumar. "Anomaly Detection for Discrete Sequences: A Survey", *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, n.24, v.5, pp. 823–839, 2009.
- [11] W. Cui e H. Wang. "Anomaly Detection and Visualization of School Electricity Consumption Data", *International Conference on Big Data Analysis*, pp. 606–611, 2017.
- [12] X. Li, Z. Li, J. Han, Jae-Gil Lee. "Temporal Outlier Detection in Vehicle Traffic Data", *IEEE International Conference on Data Engineering*, pp. 1319–1322, 2009.
- [13] Y. Heo, R. Choudhary e G. Augenbroe. "Calibration of building energy models for retrofit analysis under uncertainty", *Energy and Buildings*, vol. 47, pp. 550–560, 2012.