

UMA SISTEMÁTICA PARA DETECÇÃO DE FRAUDES EM EMPRESAS DE ABASTECIMENTO DE ÁGUA

DIEGO DE CASTRO FETTERMANN, KELLY CERQUEIRA GUERRA, ALINE PATRICIA MANO e GIULIANO DE ALMEIDA MARODIN

RESUMO

A ocorrência de fraudes no sistema das empresas de abastecimento de água resulta em um importante fator de prejuízos. Como forma de contribuir para eliminar esse problema, o objetivo desse trabalho é propor uma sistemática para direcionar a inspeção dos possíveis locais com fraude. Para isso, foi realizado um estudo de caso que se iniciou com uma pesquisa bibliográfica sobre sistemas de abastecimento de água e os métodos recomendados de serem utilizados para a detecção

de fraudes: CEP, previsão de demanda e fraudes, e detecção de outliers. Em seguida é descrito o método de pesquisa, com a descrição do caso e as etapas para o desenvolvimento da sistemática. Os resultados apontam que a aplicação da sistemática proporciona uma taxa de detecção de 98% dos casos de fraude no consumo de água, se mostrando uma alternativa capaz de direcionar os locais a serem inspecionados quanto à existência de irregularidades.

 As perdas que ocorrem no sistema de abastecimento de água e de esgotamento sanitário consistem um importante problema para o setor (Gomes, 2009). Essas perdas afetam a sustentabilidade do abastecimento exercendo impacto sobre o meio ambiente em razão do aumento do volume extraído dos mananciais (Shubo, 2003). As perdas durante o abastecimento acontecem devido à ocorrência de vazamentos, erros de medição, ligações clandestinas, fraudes e falhas no cadastro comercial (Tsutiya, 2006). A dificuldade de detecção das fraudes resulta em prejuízos financeiros para a concessionária em razão do volume de água não faturado. A

dificuldade de detecção das fraudes acontece em razão das diversas fontes de variabilidade existente no padrão de consumo de uma unidade, tais como alteração na quantidade de moradores, viagens, clima, vazamentos entre outros. A partir disso, a capacidade de detecção das fraudes se constitui em importante fator de economia no processo de distribuição de água das empresas.

As empresas de abastecimento de água enfrentam dificuldade para localizar os pontos onde ocorrem fraudes e vazamentos. O principal método de detecção utilizado é a inspeção com a utilização de equipamentos acústicos, um método vago e dispendioso, que exige percorrer

toda extensão das tubulações em busca do ponto de vazamento (Gumier e Luvizotto Junior, 2007). Na maioria dos casos, a seleção dos locais para inspeção é realizada ao acaso. Mesmo quando existe a tentativa de direcionar a inspeção, a seleção das áreas é realizada empiricamente, dificultando a localização das fraudes (Gumier e Luvizotto Junior, 2007).

O objetivo principal desse trabalho é propor uma sistemática para detecção dos locais das fraudes no sistema de abastecimento de água. A partir da aplicação dessa sistemática pretende-se identificar com maior precisão casos de fraudes (violações nos hidrômetros). Sistemas de apoio para a detecção de

PALAVRAS CHAVE / Abastecimento de Água / Controle Estatístico de Processo / Detecção de Fraudes / Detecção de Outliers /

Recebido: 21/02/2014. Modificado: 19/01/2015. Aceito: 21/01/2015.

Diego de Castro Fettermann. Doutor em Engenharia de Produção, Universidade Federal do Rio Grande do Sul (UFRGS), Brasil. Professor, Universidade Federal de Santa Catarina (UFSC). Endereço: Departamento de Engenharia de Produção e Sistemas, UFSC. Caixa-postal: 476. Trindade 88040900 -Florianópolis, SC, Brasil. Brasil. e-mail: d.fettermann@ufsc.com

Kelly Cerqueira Guerra. Graduada em Engenharia de Produção, Universidade Estadual de Santa Cruz, Brasil.

Aline Patricia Mano. Mestre em Engenharia de Produção, Universidade Federal de São Carlos, Brasil. Professora, Universidade Estadual de Santa Cruz, Brasil. E-mail: alinepatricia07@hotmail.com

Giuliano de Almeida Marodin. Doutor e Pós-doutor em Engenharia de Produção, UFRGS, Brasil. e-mail: gamarodin@gmail.com

fraudes também são utilizados para identificar fraudes corporativas (Perera *et al.*, 2014), na utilização de cartões de crédito (Eliás *et al.*, 2011) e em transações realizadas pela web (Lima e Pereira, 2012). Em sistemas de abastecimento, este tema é discutido principalmente em casos associados ao setor elétrico (Muniz *et al.*, 2008; Costa *et al.*, 2013).

A efetividade da sistemática proposta é testada em uma amostra de dados de consumo mensal de água de domicílios residenciais urbanos disponibilizados pela Empresa Baiana de Águas e Saneamento S.A. (EMBASA), empresa responsável pela distribuição de água no município de Jequié/BA, localizado na região nordeste do Brasil. Iniciativas direcionadas ao controle do consumo contribuem para a redução dos problemas no abastecimento, com destaque para a região nordeste do Brasil, a mais afetada por problemas no abastecimento de água (Rebouças, 1997). A aplicabilidade da sistemática proposta está limitada ao caso de economias residenciais.

Revisão de la Literatura

Perdas no sistema de abastecimento de água

O sistema de abastecimento de água pode ser caracterizado como o conjunto de obras, equipamentos e serviços destinados ao fornecimento de água potável e em quantidade suficiente para uma comunidade com a finalidade de consumo doméstico, industrial, em serviços públicos entre outros usos (Azevedo Netto *et al.*, 1998). As perdas podem ocorrer em todos os pontos de um sistema de abastecimento, desde a captação até as ligações prediais (Gomes, 2009). No Brasil, o índice de perdas no sistema de abastecimento de água é de cerca de 40% (SNIS, 2010). Esta perda de água não se resume apenas ao volume de água produzido que se perde no trajeto até o uso final pelos clientes, mas também a água entregue e por alguma ineficiência, não for faturada. (Tsutiya, 2006).

Estas perdas ainda podem ser caracterizadas como perdas aparentes e perdas reais (Gumier e Luvizotto Junior, 2007). As perdas aparentes, foco deste trabalho, estão relacionadas ao volume de água produzido e não faturado pela empresa devido a fraudes, erros de medição, ligações clandestinas, entre outros. As perdas reais (ou físicas) são ocasionadas por vazamentos, que ocorrem em todo o sistema de abastecimento (Vilanova *et al.*, 2010).

As fraudes são caracterizadas por intervenções improvisadas no

hidrômetro afim de que seja medida apenas uma parcela da água consumida. Outra situação comum são as ligações inativas, que não fazem parte do cadastro comercial, mas por ação do cliente se tornaram ativas sem o conhecimento da empresa (Tsutiya, 2006). As ligações clandestinas podem ser realizadas de tal forma que fica difícil sua detecção, exigindo que sejam realizadas análises de consumo ou pesquisas de campo para efetivamente identificar a ligação (Tsutiya, 2006). A adoção de um sistema de detecção de fraudes no abastecimento de água permite às empresas distribuidoras uma economia nos seus processos, em razão da redução do consumo de energia elétrica, além de custos operacionais de mão de obra e insumos (Gumier e Luvizotto Junior, 2007).

Métodos para a detecção de fraude

A literatura apresenta alguns métodos que poderiam ser utilizados para a detecção de fraudes no sistema de abastecimento de água. Esses métodos pretendem aumentar o foco de investigação, direcionando a aplicação dos instrumentos de inspeção em locais com maior probabilidade de existência de fraude. De forma geral, esta busca por fraudes acontece aleatoriamente nas economias, sendo que a utilização destes métodos procura retirar esta aleatoriedade e direcionar a detecção de fraudes em economias com maior probabilidade de ocorrência. Entre os métodos recomendados pela literatura para este objetivo o mais mencionado é o 'Controle Estatístico de Processo' (CEP; Johnson, 1984; Bolton e Hand, 2002; Morais; Pacheco, 2010). Esta ferramenta procura identificar causas especiais de variação de um processo (Montgomery, 2004; Ritzman e Krajewski, 2004; Slack *et al.*, 2008), sendo que esta causa poderia ser interpretada como um potencial problema (vazamento/fraude). Outro método mencionado são as técnicas de 'previsão de demanda' (Souza *et al.*, 2012). Essas técnicas se constituem em um conjunto de procedimentos fundamentados em modelos estatísticos, matemáticos, econométricos ou subjetivos a fim de determinar dados futuros (Martins e Laugeni, 2012). Para os casos em que o consumo de água da economia estiver abaixo ou acima do erro admissível para a previsão poderia identificar a existência de fraude no sistema. Outro método mencionado na literatura consiste em metodologias para a identificação de dados atípicos (*outliers*) (Knorr *et al.*, 1999; Ben-Gal, 2005; Araújo, 2010; Pais, 2011; Silva, 2011). Um *outlier* consiste em uma observação que se desvia muito das demais observações, a ponto de suspeitar-se

que tenha sido gerada por um mecanismo diferenciado (Hawkins, 1980), neste caso a fraude. A partir disso, a revisão de literatura está baseada nesses três métodos recomendados para a detecção de fraudes.

Controle estatístico de processo

O controle estatístico de processo (CEP) é recomendado para controlar e melhorar a qualidade de bens e serviços e eliminar a variabilidade existente em um processo por meio da utilização de técnicas estatísticas e atividades de coleta, análise e interpretação de dados (Siqueira, 1997; Ritzman e Krajewski, 2004). Existem dois tipos de causas para as variações de um processo: as causas comuns e as causas especiais ou atribuíveis. As causas comuns são as fontes de variações aleatórias e inerentes ao processo. As causas especiais compreendem qualquer fonte de variação que pode ser identificada e eliminada (Montgomery, 2004; Ritzman e Krajewski, 2004; Slack *et al.*, 2008). Uma das principais ferramentas do CEP é o gráfico de controle, sendo utilizado para o monitoramento de processos e aplicado para identificar o tipo de variabilidade presente (Montgomery, 2004; Siqueira, 2004).

Existem muitas aplicações bem sucedidas da aplicação do CEP no ambiente de fabricação, além de diversas aplicações no ambiente não industrial e em serviços. As aplicações fora da indústria exigem maior flexibilidade no uso em razão de que os sistemas não industriais tendem a possuir uma percepção mais complexa de seu funcionamento (Montgomery, 2004). Os gráficos de controle estão entre os dispositivos utilizados na estatística e suas aplicações atuais incluem administração, clínica química, epidemiologia, detecção de fraudes, cuidados de saúde, segurança, gestão de pessoal, monitoramento de resíduos, atletismo, genética, finanças, entre outras aplicações (Morais e Pacheco, 2010). A partir disso, o emprego dos gráficos de controle na detecção de fraudes, fornece técnicas eficazes capazes de detectar com sucesso atividades como lavagem de dinheiro, fraude em cartão de crédito, fraude nas telecomunicações, invasão de computador e fraude médica e científica (Bolton e Hand, 2002).

Previsão de demanda

Os métodos de previsão de demanda consideram que os mesmos fatores que influenciaram a demanda no passado, continuarão presentes no futuro (Corrêa e Corrêa, 2004). A adequação dos métodos de previsão existentes depende

de fatores como: disponibilidade de dados, tempo, recursos e horizonte de previsão (Moreira, 2008).

As técnicas de previsão podem ser divididas em duas abordagens: qualitativa e quantitativa (Corrêa e Corrêa, 2013). A qualitativa valoriza os dados subjetivos e é baseada no julgamento e intuição de especialistas, podendo ser empregada quando os fatores subjetivos são capazes de esclarecer o futuro. Os principais métodos qualitativos são: método Delphi, júri de executivos, força de vendas, pesquisa de mercado e analogia histórica (Corrêa e Corrêa, 2013; Tubino, 2008). A abordagem quantitativa é caracterizada pela necessidade de uma série de dados históricos, através dos quais são identificados padrões no comportamento da demanda que serão projetados para o futuro (Corrêa e Corrêa, 2013). Os principais métodos quantitativos são: médias móveis, suavização exponencial, Box-Jenkins, regressão linear e redes neurais (Pellegriani e Fogliatto, 2001; Corrêa e Corrêa, 2013). Apesar de possuírem boa capacidade de previsão, as técnicas quantitativas têm por características possuírem a necessidade de uma série de dados históricos. A partir do emprego das técnicas de previsão de demanda nesse trabalho, procura-se identificar um consumo esperado para cada economia. Quando os valores do consumo efetivo não estiverem dentro da margem de erro proposta pelo modelo de previsão adotado, este será considerado um indício de fraude no sistema.

Deteção de outliers

Em alguns conjuntos de dados é possível perceber que existem elementos que se distanciam significativamente dos demais, sendo comumente designados de *outliers* (Muñoz-García *et al.*, 1990). Verifica-se entre os autores que os *outliers* são dados diferenciados do conjunto geral de dados, podendo inclusive ser gerados por mecanismos diferentes (Hawkins, 1980; Liu, 2004). No caso desse estudo este mecanismo seria a fraude.

Os *outliers* se relacionam com os outros elementos da amostra e aparecem em uma quantidade pequena, pois se existissem muitos *outliers* seria impossível estabelecer um padrão (Silva, 2004). As causas do surgimento dos *outliers* podem ajudar a definir a forma como devem ser tratados. Independente de suas causas, a fase inicial do estudo de *outliers* é a deteção das observações discrepantes. A escolha do método utilizado para a deteção de *outliers* depende de diversos fatores, como a distribuição dos dados, se os parâmetros de distribuição

(por exemplo, média e variância) são conhecidos, o número esperado de dados discrepantes e o tipo de *outliers* esperados (por exemplo, superior ou inferior) (Knorr *et al.*, 1999).

Os métodos de deteção de *outliers* podem ser divididos em métodos univariados e métodos multivariados. Entre os métodos univariados mais utilizados está o Z-score com ou sem ajuste de Bonferroni (Cousineau e Chartier, 2010). Além desse, também são recomendados o teste box plot, teste do Z-score modificado, teste de Grubbs, teste de Dixon, teste de Cochran e correlação linear. Atualmente, a deteção de *outliers* é amplamente utilizada e pesquisada por diversas áreas (Araújo, 2010). Dentre as diversas aplicações da deteção de *outliers* pode-se citar: deteção de fraude, processamento de pedido de empréstimo, deteção de invasão de rede, diagnósticos de falhas, monitoramento de condições médicas, análise de irregularidade em votações, ensaios clínicos, previsão de mau tempo, análise de desempenho de atletas (Ben-Gal, 2005; Araújo, 2010).

Proposta da Sistemática para a Deteção de Fraudes

Descrição do caso

A aplicação da sistemática proposta foi testada para a deteção de fraudes na distribuição de água da Empresa Baiana de Águas e Saneamento S.A. (EMBASA), na unidade de Jequié/BA, situada na região nordeste do Brasil. Essa unidade atende um volume de 53000 economias com um volume de água tratada e distribuída de 22100m³/dia. A deteção de fraudes no sistema de abastecimento é um problema na EMBASA/Jequié, sendo que o procedimento de fiscalização realiza a seleção das economias de forma aleatória, tornando a identificação de fraudes uma tarefa pouco eficiente e onerosa à empresa. A sistemática proposta procura identificar casos em que foi constatada a presença de fraude na economia além de não acusar a presença de fraude nas demais economias.

Etapas da sistemática

A proposta de sistemática para deteção de fraudes no sistema de abastecimento de água está baseada em cinco etapas. A proposição destas etapas está relacionada com a necessidade identificada para a resolução do problema (Tabela I). Cada uma das etapas possui uma função específica e procura atender a requisitos necessários para aplicar os métodos propostos e analisar a sua

capacidade de detectar fraudes no sistema de abastecimento de água. A partir da elaboração da sistemática para deteção de fraudes, é desenvolvido um roteiro para a aplicação de cada uma das etapas propostas. Este roteiro apresenta uma descrição do objetivo da etapa, as informações de entrada utilizadas, as ferramentas ou técnicas sugeridas para serem aplicadas na etapa e as informações de saída de cada etapa (Tabela I).

Os resultados seguem as cinco etapas propostas para a aplicação da sistemática de deteção de fraudes apresentada na seção anterior.

Análise descritiva dos dados

A EMBASA disponibilizou uma série histórica de consumo de água de 13 períodos (meses), compreendidos entre julho 2012 e julho 2013, de 55 economias residenciais. Nesta amostra, foram disponibilizados 50 casos aleatórios de economias residenciais situadas no município de Jequié/BA e cinco casos em que foi detectada fraude no sistema de abastecimento de água, totalizando o conjunto de dados de consumo de 55 economias. A (Tabela II). apresenta a análise descritiva do consumo de água mensal da amostra das economias residenciais analisadas.

Seleção do método para deteção de fraudes

Após a análise exploratória dos dados foi possível identificar uma grande variabilidade nos dados de consumo de cada economia. Essa variabilidade é decorrente de causas inerentes ao processo, tais como alteração da quantidade de pessoas na residência, sazonalidade do consumo e clima, entre outros. A utilização da técnica de controle estatístico de processo é recomendada para um sistema estável, que possua sua variabilidade resultado prioritariamente de causas naturais do processo (Montgomery e Runger, 2003). No caso do consumo de água de cada economia analisada, a presença de diversas fontes de variabilidade no consumo, além das potenciais fraudes, dificultam a utilização desta técnica. Dentre os motivos que afetam o consumo de água pelas economias podem ser mencionados o clima, fluxo de visitantes e hábitos da população, entre outros (Yassuda *et al.*, 1976). A literatura recomenda a utilização de métodos quantitativos, visto que são menos tendenciosos e fazem um uso mais eficiente dos dados (Armstrong, 2001). O efeito desses motivos no consumo de água possuem uma forte componente

TABELA I
PROPOSTA DE SISTEMÁTICA PARA A DETECÇÃO DE FRAUDES, IDENTIFICANDO INFORMAÇÕES DE ENTRADA, SAÍDA E FERRAMENTAS RECOMENDADAS

Necessidade	Etapa	Descritivo	Entrada	Ferramenta	Saída
Identificar o comportamento dos dados e suas características	1. Análise Descritiva dos dados	Visa compreender e interpretar características, tais como medidas de tendência central e dispersão, além de verificar o comportamento da série temporal.	Conjunto de dados de consumo	Análise da tendência central (média e mediana), análise da dispersão dos dados, análise do coeficiente de variação (CV), análise da série histórica dos dados.	Análise descritiva dos dados
Localizar o método mais adequado para identificar fraudes no caso analisado	2. Seleção do método para detecção de fraudes	Procura-se verificar, a partir das suposições de aplicação dos três métodos, o mais adequado para detecção de fraudes no caso analisado.	Análise descritiva dos dados de consumo	Análise das suposições para aplicação dos métodos.	Método selecionado para detecção de fraudes
Adequar o comportamento dos dados selecionados para viabilizar a aplicação do método escolhido	3. Tratamento dos dados	Verificar necessidade de algum tipo de transformação nos dados ou redução de sua variabilidade para o melhor desempenho da técnica selecionada.	Conjunto de dados de consumo	Transformação dos dados, análise de cluster, testes de aderência à distribuição normal (KS, Shapiro-Wilks, Anderson Darling) ou outra distribuição contínua de probabilidade.	Conjunto de dados de consumo tratados e analisados
Realizar aplicação do método de detecção de fraudes escolhido	4. Aplicação do método selecionado	Aplicar o método selecionado seguindo as recomendações propostas pela bibliografia.	Conjunto de dados de consumo tratados e analisados	Aplicação do método selecionado (CEP, previsão de demanda e detecção de outliers).	Resultados da aplicação do método
Analisar a capacidade em detectar fraudes no caso estudado	5. Análise dos resultados	Identificar a taxa de acerto na identificação das economias com fraude e a taxa de indicação de fraude em casos em que não foi detectado (falso positivo).	Resultados da aplicação do método	Taxa de 'erro tipo I', taxa de 'erro tipo II'.	Análise dos resultados

TABELA II
ANÁLISE DESCRITIVA DO CONSUMO DE ÁGUA MENSAL

	Economias sem fraude (n=50)			Economias com fraude (n=5)		
	Consumo médio (m³)	Desvio padrão	Coefficiente de variação	Consumo médio (m³)	Desvio padrão	Coefficiente de variação
Jul/2012	28,68	19,18	0,67	9,0	8,22	0,91
Ago/2012	28,44	17,75	0,62	7,4	7,40	1,00
Set/2012	28,60	15,43	0,54	2,2	2,17	0,99
Out/2012	24,50	14,38	0,59	1,8	1,64	0,91
Nov/2012	27,16	13,03	0,48	2,0	1,22	0,61
Dez/2012	27,60	14,11	0,51	3,4	3,21	0,94
Jan/2013	29,68	13,73	0,46	0,6	1,34	2,23
Fev/2013	34,12	19,33	0,57	1,4	1,34	0,96
Mar/2013	26,28	18,03	0,69	5,4	7,40	1,37
Abr/2013	25,92	14,93	0,58	6,2	5,72	0,92
Mai/2013	27,02	12,63	0,47	8,2	6,72	0,82
Jun/2013	25,90	11,56	0,45	11,6	5,68	0,49
Jul/2013	31,74	13,58	0,43	14,4	7,80	0,54
Média	28,13	15,21	0,54	5,66	4,60	0,98

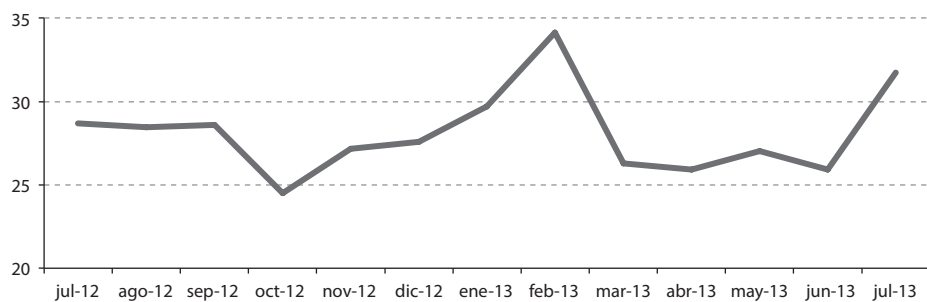


Figura 1. Consumo médio mensal da amostra.

sazonal. Para a modelagem deste componente pela série temporal é necessária uma maior quantidade de períodos que os disponíveis (13 períodos), fator este que prejudica a utilização de técnicas de previsão de séries temporais quantitativas com tendência sazonal.

Em razão dessas dificuldades, foi selecionado o método de detecção de *outliers*. Existem diversos métodos para a detecção de *outliers* univariados, como no caso deste estudo. Entre diversas opções, a utilização da técnica de Z-score modificado obteve melhor desempenho em um caso para detecção de fraudes em energia elétrica (Nascimento *et al.*, 2012). Cousineau e Chartier (2010) também mencionam a utilização do Z-score como um dos mais utilizados. Os métodos escolhidos foram o Z-Score e o Z-Score modificado, em razão de sua eficácia e ampla aceitação (Shiffler, 1988; Crosby, 1994; Seo, 2006), sendo que seus resultados para adeteção de fraudes serão comparados.

Tratamento dos dados

A grande variabilidade dos dados de consumo de água das economias da amostra é explicada, em parte, pela variação entre essas economias. Dentro da amostra aleatória de 50 economias residenciais, existem casos de

residências grandes, com diversos pontos de água e residências pequenas com menor quantidade de pontos e consumo. Para reduzir a variabilidade de dados de consumo utilizou-se a técnica multivariada análise de *cluster*, que tem por objetivo identificar agrupamentos com semelhança interna e heterogêneos entre si (Tinsley e Brown, 2000; Hair *et al.*, 2005; Norusis, 2011). Foi analisada a série de 13 meses de consumo de água do conjunto de 50 economias da amostra por meio da análise de *cluster* hierárquico, seguindo o método de Ward e o algoritmo do quadrado da distância euclidiana (Figura 2).

O resultado obtido indica a presença de dois grandes agrupamentos, o primeiro agrupamento inclui economias com padrão maior de consumo de água na amostra e o segundo com economias de menor consumo (Tabela III). Os resultados indicam uma redução da variabilidade do consumo de água entre os grupos identificados.

A partir disso, foram realizados testes para verificar a aderência dos dados à distribuição normal. Com este objetivo foi utilizado o testes Kolmogorov Smirnov (KS), sendo que os resultados para todos os períodos aceitam a hipótese de aderência à distribuição normal ($p > 0,005$).

HIERARCHICAL CLUSTER ANALYSIS

Dendrogram using Ward Method

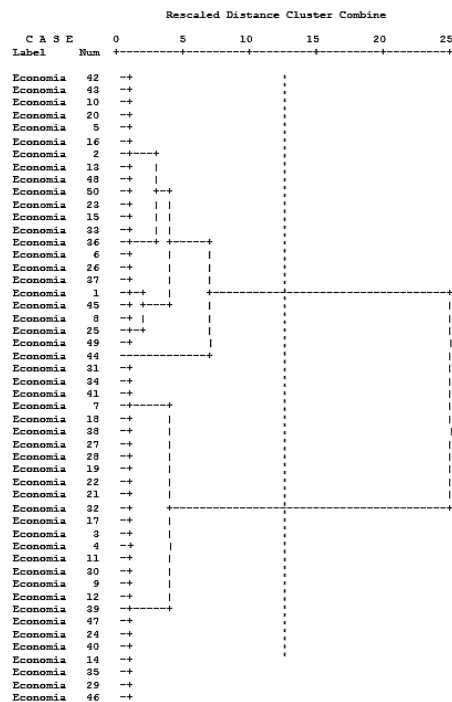


Figura 2. Dendrograma da análise de cluster hierárquico. Fonte: SPSS® v 17.

Aplicação do método selecionado

O Z-score é um método simples utilizado em dados com distribuição normal que identifica *outliers* utilizando a média e o desvio padrão da amostra (Shiffler, 1988; Seo, 2006). O Z-score é calculado através da Eq. 1 (Shiffler, 1988):

$$Z_{score} = \frac{x_i - \bar{x}}{s} \quad (1)$$

onde

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

É sugerido que um dado seja considerado *outlier* quando o valor absoluto do Z-score for >3 , no entanto, a depender das características da amostra ou dos critérios de decisão especificados, este valor pode ser alterado (Nascimento *et al.*, 2012). Uma limitação para o uso do Z-score é o desvio padrão elevado em razão dos dados com valores extremos. Dessa forma, os *outliers* menos extremos passam despercebidos por causa dos *outliers* mais extremos (Seo, 2006). Para evitar esse problema, foi desenvolvido o método do Z-score modificado (Seo, 2006; Nascimento *et al.*, 2012). A média e o desvio padrão são substituídos por indicadores mais robustos: a mediana e a média do desvio absoluto da mediana (MAD) (Eq. 3).

$$MAD = \text{median} \{ |x_i - \bar{x}_i| \} \quad (3)$$

A utilização desses indicadores resulta no método Z-score modificado (Seo, 2006; Nascimento *et al.*, 2012). Sendo assim, o Z-score modificado (M_i) é calculado a partir da mediana e do

desvio absoluto da mediana (MAD), além de uma constante (Eq. 4) (Shiffler, 1988). São considerados *outliers* as observações com $M_i < K$, sendo K um valor selecionado pelo pesquisador.

$$M_i = \frac{0,6745}{MAD} (x_i - \bar{x}_i) \quad (4)$$

Utilizando o software Excel® foi calculado o Z-score e o Z-score modificado para cada residência do Grupo 2 e para as residências com fraude já confirmada durante os treze períodos. As observações pertencentes ao Grupo 1 não foram envolvidas, pois se tratam de valores de consumo superiores e dificilmente possuam fraude. Foram considerados como fraude os casos com identificação de *outliers* por, pelo menos, três períodos (meses) consecutivos.

No cálculo do Z-score foi considerado $K = -1,5$ e pode-se perceber que, dentre as cinco economias com fraude confirmada o método foi capaz de identificar quatro delas e, no grupo das demais economias, duas foram identificadas como possíveis fraudes. Para o Z-score modificado foi utilizado $K = -1,3$, sendo possível observar que todas as economias com fraude confirmada foram identificadas como tal e, nas demais economias duas foram apontadas como possível fraude.

Análise dos Resultados

Ao testar qualquer hipótese existem dois tipos de erros que podem surgir. O ‘erro tipo I’ é a rejeição de uma hipótese nula (H_0) quando ela é, de fato, verdadeira. O ‘erro tipo II’ consiste em não rejeitar a hipótese nula quando,

TABELA III
ANÁLISE DESCRITIVA DO CONSUMO DE ÁGUA MENSAL DOS GRUPOS 1 E 2

	GRUPO 1 (n=23)			GRUPO 2 (n=27)		
	Consumo médio (m³)	Desvio padrão	Coefficiente de variação	Consumo médio (m³)	Desvio padrão	Coefficiente de variação
Jul/2012	43,52	18,27	0,42	16,03	7,07	0,44
Ago/2012	41,48	16,48	0,40	17,33	9,24	0,53
Set/2012	39,65	13,44	0,34	19,18	9,78	0,51
Out/2012	35,48	11,11	0,31	15,15	9,39	0,62
Nov/2012	37,56	9,09	0,24	18,30	8,51	0,47
Dez/2012	39,04	10,31	0,26	17,85	8,40	0,47
Jan/2013	40,96	8,78	0,21	20,07	9,05	0,45
Fev/2013	49,43	15,56	0,31	21,07	10,74	0,51
Mar/2013	37,78	19,96	0,53	16,48	7,68	0,47
Abr/2013	37,09	14,60	0,39	16,41	5,94	0,36
Mai/2013	36,35	11,83	0,33	19,07	6,32	0,33
Jun/2013	33,78	11,21	0,33	19,18	6,63	0,35
Jul/2013	36,43	11,64	0,32	27,74	14,03	0,51
Média	39,12	13,25	0,34	18,76	8,68	0,46

na verdade, ela é falsa (Fonseca; Martins, 2006). No caso de incidir em um erro do tipo I, se indica um caso de fraude sendo que esta não existe. No caso de incidência de erro tipo II, os casos de fraude não são identificados pelo sistema. Tanto a incidência dos erros do tipo I quanto do tipo II tende a dificultar o processo de inspeção de fraudes. O primeiro tende a sobrecarregar o trabalho das equipes de inspeção, enquanto que o segundo não indica casos de fraude quando eles realmente existem. A incidência dos Erros nos métodos utilizados é apresentada na (Tabela IV).

Os resultados encontrados indicam uma maior taxa de acerto do método Z-score modificado. Mesmo assim, ainda não se pode afirmar que nas economias em que foram identificadas fraudes (erro tipo I) não exista fraude, visto que essas economias foram selecionadas aleatoriamente do sistema de clientes da empresa e não foram inspecionadas para a localização de fraudes.

Conclusões

O estudo realizado propõe uma sistemática para detecção dos locais das fraudes no sistema de abastecimento de água. O resultado obtido permite direcionar um maior foco ao processo de inspeção de fraudes no sistema de abastecimento. Apesar de a taxa de erro obtido pelo método de Z-score modificado ser considerada ainda alta (2/55), os resultados permitem direcionar o processo de inspeção nas unidades realizado aleatoriamente pela empresa.

Apesar da utilização da técnica de detecção de *outliers*, ressalta-se a necessidade de contornar as restrições indicadas para a utilização das demais técnicas de detecção de fraudes sugeridas pela literatura. Como forma de melhorar os resultados encontrados, indica-se: i) a disponibilidade de algum critério de classificação das economias em relação ao número de moradores ou à quantidade de pontos de água a fim de reduzir a variabilidade na demanda; ii) a disponibilização de uma série maior de dados, que permitiria explorar a utilização de técnicas quantitativas de previsão de demanda para a identificação de fraudes; iii) analisar mais profundamente fatores de variação no consumo de forma a identificar fraudes como causas

especiais; e iv) operacionalização da sistemática de detecção de fraudes proposta na forma de um sistema.

REFERÊNCIAS

- Araújo BM (2010) *Identificação de outliers em redes complexas baseada em caminhada aleatória*. These. Universidade de São Paulo, Brasil. 74 pp.
- Armstrong J (2001) Standards and practices for forecasting. Em Armstrong J (Ed.) *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Kluwer. Boston, MA, EEUU. pp. 679-732.
- Azevedo Netto JM, Fernandez MF, Araujo R, Ejiito A (1998) *Manual de Hidráulica*. 8ª ed. Blucher. São Paulo, Brasil. 680 pp.
- Ben-Gal I (2005) Outlier detection. Em Maimon O, Rokach L (Eds.) *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*. Kluwer. Boston, MA, EEUU. pp. 117-130.
- Bolton RJ, Hand DJ (2002) Statistical fraud detection: A review. *Stat. Sci.* 17: 235-255.
- Corrêa HL, Corrêa CA (2013) *Administração de Produção e Operações: Manufatura e Serviços: Uma Abordagem Estratégica*. (2ª Ed.) Atlas. São Paulo, Brasil. 520 pp.
- Costa BC, Alberto BLA, Portela AM, Maduro W, Esdras OE (2013) Fraud detection in electric power distribution networks using an ann-based knowledge-discovery process. *Int. J. Artif. Intell. Applic.* 4(6): 17-23.
- Cousineau D, Chartier S (2010) Outliers detection and treatment: a review. *Int. J. Psychol. Res.* 3: 58-67.
- Crosby T (1994) How to detect and handle outliers. *Technometrics* 3: 315-316.
- Eliás, A, Zezzatii AO, Padilla A, Ponce J (2011) Outlier analysis for plastic card fraud detection: a hybridized and multi-objective approach. *Hybrid Artif. Intell. Syst.* (II): 1-9.
- Fonseca JS, Martins GA (2006) *Curso de Estatística*, (6ª ed.) Atlas. São Paulo, Brasil. 320 pp.
- Gomes HP (2009) *Eficiência Hidráulica e Energética em Saneamento: Análise Econômica de Projetos* (2ª ed.) Editora Universitária da UFPA. João Pessoa, Brasil. 145 pp.
- Gumier CC, Luvizotto Junior EJ (2007) Aplicação de modelo de simulação-otimização na gestão de perda de água em sistemas de abastecimento. *Rev. Eng. Sanit. Amb.* 12: 32-41.
- Hair JF, Anderson RE, Tatham RL, Black WC (2005) *Multivariate Data Analysis*. (5ª ed.) Prentice Hall. Upper Saddle River, NJ, EEUU. 982 pp.
- Hawkins D (1980) *Identification of Outliers*. Chapman & Hall. Londres, RU. 188 pp.
- Johnson DG (1984) "Trial by Computer": A case study of the use of simple statistical

techniques in the detection of a fraud. *J. Operat. Res. Soc.* 35: 811-820.

- Knorr EM, Ng RT, Tucakov V (1999) Distance-based outliers: algorithms and applications. *VLDB J.* 8: 237-253.
- Lima RAF, Pereira ACM (2012) Fraud detection in web transactions. Em *Proc. 18ª Braz. Symp. Multimedia and the web*. ACM: 273-280.
- Liu H, Shah S, Jiang W (2004) On-line outlier detection and data cleaning. *Comput. Chem. Eng.* 28: 1635-1647.
- Martins PG, Laugeni FP (2012) *Administração da Produção*. (2ª ed.) Saraiva. São Paulo, Brasil. 562 pp.
- Montgomery DC (2004) *Introdução ao Controle Estatístico da Qualidade* (4ª ed.) LTC, Rio de Janeiro, Brasil. 532 pp.
- Montgomery DC, Runger GC (2003) *Estatística Aplicada e Probabilidade para Engenheiros*. (2ª ed.) LTC. Rio de Janeiro, Brasil. 463 pp.
- Morais MC, Pacheco A (2010) Combined CUSUM-Shewhart schemes for binomial data. *Econ. Qual. Contr.* 21: 43-57.
- Moreira DA (2008) *Administração da Produção e Operações*. (2ª ed.) Cengage. São Paulo, Brasil. 624 pp.
- Muniz C, Figueiredo K, Vellasco M, Pacheco M, Chavez G (2008) Indicação de suspeitos de irregularidade em instalações elétricas de baixa tensão. *Learn. Nonlin. Mod.* 6: 16-28.
- Muñoz-García J, Moreno-Rebollo JL, Pascual-Acosta A (1990) Outliers: a formal approach. *Int. Stat. Rev.* 58: 215-226.
- Nascimento RM, Oening AP, Marcilio DC, Aoki AR, Rocha JR, Schiochet JM (2012) Algoritmo de detecção e correção de outliers para previsão de carga. *Anais IV Simp. Bras. Sistemas Elétricos (SBSE)*. Goiânia, Brasil. p. 1-6.
- Norusis MJ (2011) *PASW Statistics 18: Advanced Statistical Procedures Companion*. SPSS. Chicago, IL, EEUU.
- Pais RMA (2011) *Detection of Outliers and Outliers Clustering on Large Datasets with Distributed Computing*. These. Universidade de Lisboa. Portugal. 188 pp.
- Pellegrini FR, Fogliatto FS (2001) Passos para implantação de sistemas de previsão de demanda: Técnicas e estudo de caso. *Produção* 11: 43-64.
- Perera LCJ, De Freitas EC, Imoniana JO (2014) Avaliação do sistema de combate às fraudes corporativas no Brasil. *Rev. Contemp. Contab.* 11(23): 3-30.
- Rebouças AC (1997) Água na região Nordeste: desperdício e escassez. *Estud. Avanç.* 11(29): 127-154.
- Ritzman LP, Krajewski LJ (2004) *Administração da Produção e Operações* (1ª ed.) Pearson. São Paulo, Brasil. 431 pp.
- Seo S (2006) *A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets*. These. University of Pittsburgh. EEUU. 53 pp.
- Shiffler RE (1988) Maximum Z scores and outliers. *Am. Stat.* 42: 79-80.
- Shubo T (2003) *Sustentabilidade do Abastecimento e da Qualidade da Água Potável Urbana*. Escola Nacional de Saúde Pública. Fund. Oswaldo Cruz. Rio de Janeiro, Brasil. 113 pp.

TABELA IV
TAXAS DO ERRO TIPO I E DO ERRO TIPO II

Método	Taxa de 'erro tipo I'	Taxa de 'erro tipo II'	Taxa de erro/acerto
Z-score	2/50	1/5	3/55
Z-score modificado	2/50	0/5	2/55

- Silva FR (2004) *Uma Abordagem para Detecção de Outliers em Dados Categóricos*. These. UNICAMP, Brasil. 81 pp.
- Silva MD (2011) *Um Modelo para Análise e Tratamento de Dados de Demanda de Energia Elétrica*. These. Universidade Federal de Alfenas. Brasil. 111 pp.
- Siqueira LGP (1997) *Controle Estatístico do Processo* (1ª ed.) Pioneira. São Paulo, Brasil. 129 pp.
- Slack N, Chambers S, Johnston R (2008) *Administração da Produção* (2ª ed.) Atlas. São Paulo. 747 pp.
- SNIS (2010) *Diagnóstico dos Serviços de Água e Esgotos-2010*. Sistema Nacional de Informações sobre Saneamento. Brasil. <http://www.snis.gov.br/PaginaCarrega.php?EWRErterterTERTer=95> (Cons. 10/06/2013).
- Souza DP, Santos RK, Santos RF (2012) Estimativa do consumo de água em restaurantes na cidade de Cascavel/PR. *Acta Iguazu* 1(3): 50-63.
- Tinsley HEA, Brown SD (2000) *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. Academic Press. San Diego, CA, EEUU. 696 pp.
- Tsutiya MT (2006) *Abastecimento de Água* (3ª ed.) Escola Politécnica, Universidade de São Paulo. Brasil. 644 pp.
- Tubino DF (2008) *Planejamento e Controle da Produção: Teoria e Prática* (1ª ed.) Atlas. São Paulo, Brasil. 208 pp.
- Vilanova MRN, Viana ANC, Salvador E, Santos GC, Pascoa C (2010) Redução do consumo de energia elétrica em sistemas de abastecimento através da redução das perdas de água: estudo de caso no município de Lavras/MG. Em Gomes HP *Sistemas de Saneamento: Eficiência Energética* (1ª ed.) Editora Universitária da UFPB. João Pessoa, Brasil. 359-366 pp.
- Yassuda ER, Oliveira WE, Gaglianone S, Nogami PS, Pereira BEB, Martins JA (1976) *Técnica de Abastecimento e Tratamento de Água*. Vol. 1. (2ª ed.) CETESB. São Paulo, Brasil. 274 pp.

A METHOD FOR FRAUD DETECTION IN WATER SUPPLY SYSTEM

Diego de Castro Fettermann, Kelly Cerqueira Guerra, Aline Patricia Mano and Giuliano de Almeida Marodin

SUMMARY

Fraud in the water supply system is an important factor of waste to the sector. To manage this problem, this paper proposes a method for fraud detection in water supply systems. This method is based in a literature review about water supply systems and techniques of fraud detection, like statistical process control (SPC), forecasting methods and outliers detection. The method proposed is applied in a practical case, with a sample from 55 households. The results appointed fraud cases with a 98% accuracy rate. The method application can help water supply system companies to guide the inspection in households' water supply.

UN MÉTODO PARA LA DETECCIÓN DE FRAUDE EN EL SISTEMA DE ABASTECIMIENTO DE AGUA

Diego de Castro Fettermann, Kelly Cerqueira Guerra, Aline Patricia Mano y Giuliano de Almeida Marodin

RESUMEN

El fraude en el sistema de abastecimiento de agua se traduce en un factor importante de pérdida para las empresas. Para manejar este problema, este trabajo propone un método para detección de fraude para el abastecimiento de agua. Este método está basado en una revisión de la literatura sobre los sistemas de abastecimiento de agua y técnicas para la detección de fraudes, tales como el control estadístico de procesos (CEP), métodos de previsión y detección de datos atípicos. El método propuesto ha sido aplicado en un caso práctico, con una muestra de 55 hogares. Los resultados apuntaron los casos de fraude con una precisión del 98%. La aplicación del método puede ayudar las empresas de abastecimiento de agua a fin de orientar la inspección de los hogares en busca de fraudes.