

DETECÇÃO E CORREÇÃO AUTOMÁTICA DE *OUTLIERS* PARA SISTEMAS SUPERVISÓRIOS EM SUBESTAÇÕES

PEDRO H. M. DE ANDRADE*, JUAN M. M. VILLANUEVA†, HELON D. M. BRAZ‡

**Departamento de Engenharia Elétrica, UFPB, João Pessoa/Brasil
Cidade Universitária, João Pessoa-PB, CEP:58.051-900*

†*Departamento de Engenharia Elétrica, UFPB, João Pessoa/Brasil
Cidade Universitária, João Pessoa-PB, CEP:58.051-900*

‡*Departamento de Engenharia Elétrica, UFPB, João Pessoa/Brasil
Cidade Universitária, João Pessoa-PB, CEP:58.051-900*

Emails: `pedro.andrade@cear.ufpb.br`, `jmauricio@cear.ufpb.br`, `helon@cear.ufpb.br`

Abstract— The load data acquired at the substation metering points by power utilities are often affected by erroneous measurements, known as outliers, which affect the studies of the utilities on their distribution areas. The Supervisory System may have atypical value alerts, however, due to the high number of outliers at various measurement points, such data is not immediately addressed and corrected, thus large erroneous data sets persist. For this, a complementary module was developed based on Artificial Intelligence techniques and autoregressive models (ARIMA) for the Supervisory System that has the functions of detecting and correcting outliers. This module also proposes the use of an occurrence factor to indicate the portion of erroneous samples of the total data set. The developed algorithm was applied on real data of a substation of the distribution system for a period of six years.

Keywords— Outliers correction, load curves, Supervisory System, SCADA.

Resumo— Os dados de demanda obtidos nos pontos de medição das subestações frequentemente são afetados por medições errôneas, denominadas *outliers*, que afetam os estudos das concessionárias sobre suas áreas de distribuição. O Sistema Supervisório pode ter alertas de valores atípicos, entretanto, devido à alta quantidade de *outliers* em diversos pontos de medição esses dados não são tratados e corrigidos de maneira imediata, persistindo-se assim grandes conjuntos de dados errôneos. Para isso foi desenvolvido um módulo complementar baseado em técnicas de Inteligência Artificial e em modelos autorregressivos do tipo ARIMA para o Sistema Supervisório que possui as funcionalidades de detectar e corrigir os *outliers*. Neste módulo também se propõe o uso de um fator de ocorrências para indicar a parcela de amostras errôneas do conjunto total dos dados. O algoritmo desenvolvido foi aplicado sobre dados reais de uma subestação do sistema de distribuição para um período de seis anos.

Palavras-chave— Correção de *Outliers*, Curvas de Demanda, Sistema Supervisório, SCADA.

1 Introdução

As Redes Elétricas Inteligentes, conhecidas na literatura como *Smart Grids*, são os sistemas elétricos de potência que foram dotados de recursos de tecnologia da informação e de elevado grau de automação para melhorar significativamente a sua eficiência operacional (Bikmetova et al., 2017). De acordo com (Katic, 2016), o conceito de inteligência é agregado às diferentes áreas do sistema elétrico, pois inclui a (i) Geração Inteligente, (ii) Medição, (iii) Transmissão e Distribuição, (iv) Consumo e (v) os dispositivos inteligentes.

Para integrar todas as áreas das redes inteligentes é necessário um Centro de Controle que inclui, dentre outros campos: Sistema de Gerenciamento de Energia; Sistema de Gerenciamento de Falhas; Análise de Operações, Estimacões e Planejamento; e Sistema Supervisório. E para garantir características como alta confiabilidade, segurança, flexibilidade e estabilidade; são necessários sensores, atuadores, processamento em tempo real, novas interfaces e protocolos de comunicação que permitam a auto-recuperação do sistema elétrico de potência (H. He, 2016).

Um componente central da Rede Elétrica Inteligente é o Sistema SCADA (do inglês, *Supervisory Control and Data Acquisition*) que, diferentemente de outras redes cibernéticas, possui capacidade de lidar com componentes heterogêneos como os servidores de controle, terminais remotos e dispositivos elétricos inteligentes como sensores e atuadores. O SCADA é a principal plataforma que conecta estações geradoras, subestações, centrais de controle e centros de carga em um sistema elétrico de potência. O sistema é o principal responsável por monitorar e controlar remotamente os equipamentos na obtenção de dados e analisá-los nas centrais de comando (Georgescu, 2013).

Os medidores das subestações permitem a captura de dados em intervalos predeterminados e tais medições são armazenadas em bancos de dados de interesse da concessionária de energia (M. R. Müller, 2014). Contudo, os dados de demanda podem ser afetados por queda de energia, instabilidade do sensor, manobras de chaves ou erros de comunicação. Quando esses eventos ocorrem podem ser gerados valores atípicos desassociados do comportamento padrão da curva. Esses valores discrepantes são denominados na literatura

de *outliers* (D. C. R. Melo, 2014).

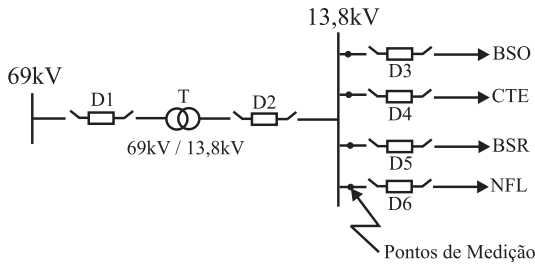


Figura 1: Planta simplificada da Subestação Estudada.

A concessionária de energia tem bastante interesse nos dados coletados nas subestações em suas áreas de distribuição, pois a partir deles podem ser desenvolvidos estudos de planejamento e operação, além do desenvolvimento de estimativas da demanda a médio e longo prazos para uma melhor alocação de recursos (Medeiros., 2016). Portanto, a qualidade dos dados coletados deve ser um aspecto buscado para o histórico armazenado, pois para realização otimizada de estudos de previsão de séries temporais é importante uma integralidade mínima dos dados; sob o risco de obter resultados equivocados ou inconclusivos (Mohammadreza et al., 2015).

A exatidão e a integralidade nos perfis de carga medidos devem ser buscadas para as atividades de previsão, planejamento e operação no setor elétrico. Além disso, notificações devem ser geradas quando *outliers* são detectados. Diante desta conjuntura, tem-se a motivação para se adotar metodologias que detectem, corrijam e indiquem a presença de *outliers* nos dados de potência de uma subestação.

Neste trabalho, é apresentado um módulo complementar ao sistema SCADA responsável pela correção das medições de potência cujos valores foram zero. Após intervalos fixos de tempo configurados pelos usuários são gerados relatórios com o resumo do número de *outliers* identificados em cada alimentador.

A abordagem para a correção desses valores atípicos foi baseada em técnicas de Inteligência Artificial (Lógica *Fuzzy* e Redes Neurais Artificiais) e em modelos autoregressivos do tipo ARIMA, utilizando os dados passados dos perfis de carga. São apresentados os resultados da correção de *outliers* considerando medições dos quatro alimentadores de uma subestação real de 69kV/13,8kV localizada no estado da Paraíba, ilustrada na Figura 1.

2 Sistema SCADA e *Outliers*

A rede de distribuição de energia é um sistema complexo em que a eficiência e a confiabilidade da operação são grandes interesses da concessionária;

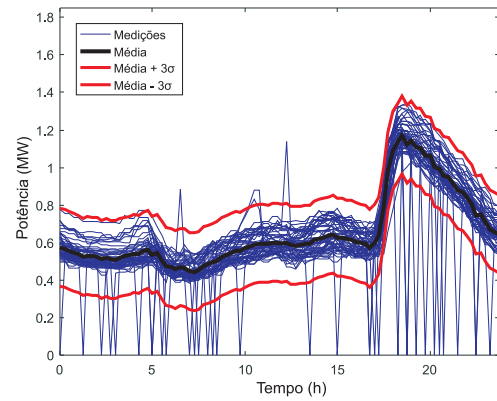


Figura 2: Curvas de todas as quartas-feiras do ano de 2010 no Alimentador 21L3-BSR da subestação estudada.

e uma das ferramentas para garantir estas características é a automação da subestação. A automação se refere a aquisição dos dados dos medidores eletrônicos inteligentes, controle e automação das manobras a partir de comandos locais ou remotos oriundos do Centro de Controle (Ramesh, 2015).

O Sistema SCADA adquire e fornece os dados em tempo real, construindo um histórico de dados para o Gerenciamento de Energia. O banco de dados criado não é usado apenas para a previsão de carga a curto prazo, mas também para o modelamento dos perfis de carga (Han and Li, 2012). A integridade e a confiabilidade das informações sobre a rede elétrica são a base nas operações do sistema e dados errôneos podem afetar a distribuição de energia ou as infraestruturas críticas. Esses dados errôneos podem ser *outliers* decorrentes de erros de medição ou até de dados maliciosos injetados em ataques contra o sistema da concessionária (Anwar et al., 2016).

Os *outliers* podem ser classificados em dois tipos: vales (zeros) ou picos (*spikes*). Os vales ocorrem quando há a ausência de dados, seja por queda na alimentação do sistema de medição ou por dano no sensor, e os valores lidos pelo SCADA são zero. Os picos ocorrem mais raramente; por erro no sensor, o valor medido possui uma magnitude significativamente diferente dos valores padrões do comportamento esperado da curva. Na Figura 2, as curvas de todas as quartas-feiras de um ano são mostradas, sendo destacadas a curva média e as curvas da média mais e menos o dobro do desvio padrão, tomando um dos alimentadores estudados durante um ano.

Em (Alves et al., 2017) é apresentada uma abordagem híbrida para detectar os outliers em dados reais adquiridos pelo SCADA de uma subestação, em (Anwar et al., 2016) utilizaram as medições do SCADA para estimação de valores e, assim, se proteger da injeção de dados maliciosos. Para aumentar a confiabilidade das medições,

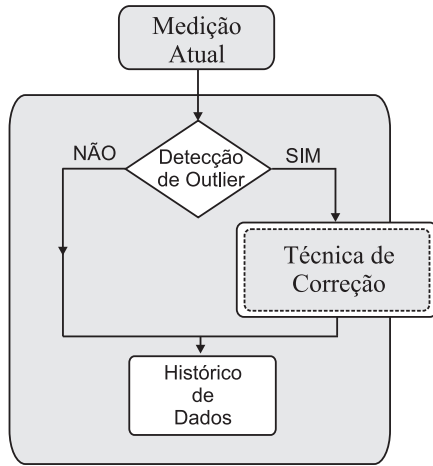


Figura 3: Fluxograma do Algoritmo de Correção de *Outliers* desenvolvido.

(Varodayan and Gao, 2010) propõem usar outro medidor para gerar redundância. Uma abordagem utilizando mineração de dados é adotada por (B. C. Neagu, 2017) para extração de indicadores das curvas e detecção dos *outliers* através de processamento estatístico.

A subestação estudada encontra-se na cidade de Cuité, no interior da Paraíba e sua topologia é de Arranjo Simples Abaixadora, como ilustrada na Figura 1. Os pontos de medição estão posicionados nos troncos dos 4 alimentadores indicados que alimentam as cidades de Baraúnas e Sossego (BSO), Cuité (CTE), Barra de Santa Rosa (BSR) e Nova Floresta (NFL).

O sistema SCADA coleta os dados da subestação de maneira sequencial a cada 15 (quinze) minutos; sendo no total 96 medições por dia. As grandezas medidas são tensão, corrente, fator de potência, potência ativa e potência reativa. As curvas de carga frequentemente possuem *outliers*, mas o sistema SCADA não é capaz de indicá-lo. Diante desse contexto, foi desenvolvido um módulo para complementar o sistema SCADA ao detectar, indicar e corrigir o *outlier*.

3 Módulo Desenvolvido para Correção dos *Outliers*

Neste trabalho, os *outliers* tratados serão os do tipo vale ou ausência de dados, isto é, quando a medição armazenada pelo SCADA tem valor igual a zero. Os *outliers* do tipo pico são bem mais raros na subestação tratada e serão desconsiderados nesta análise.

A base da abordagem do algoritmo desenvolvido consiste em fazer a previsão do próximo valor, isto é, da próxima medição realizada nos próximos 15 minutos. Nas abordagens de previsão a curtíssimo prazo (minutos ou horas à frente) devem ser levadas em consideração apenas as últimas amostras efetuadas, desconsiderando fatores como con-

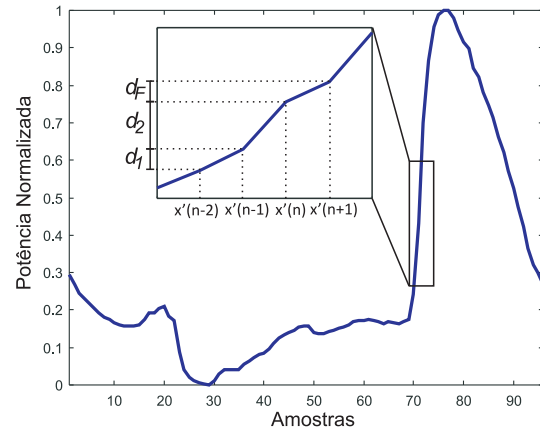


Figura 4: Gráfico com os incrementos diferenciais de entrada e de saída do Bloco *Fuzzy*.

dições climáticas e características socioeconômicas da região (Charytoniuk and Chen, 2000). A cada passo de cálculo (intervalo de 15 minutos) é verificada se a medição é um valor atípico. Em caso afirmativo, este valor será substituído por um valor calculado na iteração anterior pelo algoritmo desenvolvido. Caso negativo, a medição atual será considerada válida e integrada diretamente ao histórico de dados.

O algoritmo desenvolvido é ilustrado na Figura 3 e a técnica utilizada para correção dos *outliers* pode ser substituída. Isto é, o algoritmo é o mesmo, contudo a metodologia de correção é modular em que pode-se substituí-la de acordo com as características do perfil de carga. A seguir, serão apresentadas as três abordagens desenvolvidas: (i) Lógica *Fuzzy*, (ii) Redes Neurais Artificiais e (iii) ARIMA.

3.1 Abordagem *Fuzzy*

A Lógica *Fuzzy* é baseada na linguagem natural, isto é, na especificação de regras a partir do conhecimento do sistema pelo especialista. As relações entre as variáveis são expressas linguisticamente, proporcionando maior simplicidade e evitando um desenvolvimento matemático e estatístico mais complexo.

Para corrigir os *outliers*, inicialmente foi feita a separação dos dados históricos em sete dias da semana por alimentador no período de um ano. Foram calculados os valores máximo e mínimo de cada dia para viabilizar a normalização entre 0 e 1:

$$x'(n) = \frac{x(n) - \min(x)}{\max(x) - \min(x)} \quad (1)$$

com $x'(n)$ sendo o valor normalizado, $x(n)$ sendo a medição de entrada atual, $\max(x)$ e $\min(x)$ sendo as funções que fornecem o maior e o menor valor da série temporal naquele dia da semana da amostra em estudo. A Figura 4 ilustra as entradas do

bloco *fuzzy* que são os incrementos diferenciais dos valores anteriores dados por:

$$d_1 = x'(n-1) - x'(n-2) \quad (2)$$

$$d_2 = x'(n) - x'(n-1) \quad (3)$$

em que $x'(n-1)$ é o valor normalizado da medição imediatamente anterior, $x'(n-2)$ é o valor que precede o valor anterior e $x'(n)$ é o valor medição atual. A partir das magnitudes e dos sinais dos incrementos (d_1 e d_2) é determinada a saída do bloco (d_F) que também é um valor incremental e será somado ao valor de $x'(n)$ para determinar o valor estimado:

$$x'(n+1) = x'(n) + d_F \quad (4)$$

A cada passo de cálculo, o algoritmo identifica qual é o dia da semana, normaliza o valor de entrada de acordo com o máximo e o mínimo respectivos e calcula os incrementos diferenciais de entrada para o bloco *fuzzy*. O incremento diferencial de saída (d_F) é adicionado ao valor atual, $x'(n)$, resultando no valor estimado, $x'(n+1)$. E, por fim, é aplicada a função inversa da Eq. 1. Então se na próxima iteração houver um *outlier*, a medição será substituída por esse valor.

O comportamento dos perfis de carga foi modelado em 25 regras *fuzzy* (5 conjuntos de $d_1 \times 5$ conjuntos de d_2). E os principais pontos de análise são os que apresentam variação de sinal da derivada. Os incrementos diferenciais foram divididos de acordo com suas magnitudes e sinais em cinco conjuntos: Grande Negativo (GN), Pequeno Negativo (PN), Zero (ZE), Pequeno Positivo (PP) e Grande Positivo (GP). O incremento de saída foi construído de acordo com o comportamento do sistema: Desce Muito (DM), Desce Pouco (DP), Constante (CTE), Sobe Pouco (SP) e Sobe Muito (SM).

Através das análises sobre as curvas algumas observações podem ser feitas. Por exemplo, se as duas entradas pertencem ao conjunto Grande Negativo (GN) ou Grande Positivo (GP) indica que já houve a descida ou a subida do pico do dia, e o valor de saída deve permanecer constante. Assim como nos casos em que d_2 é zero, pois a curva tanto pode subir suavemente mesmo com d_1 negativo como pode descer suavemente com o d_1 positivo.

3.2 Abordagem RNA

A abordagem utilizando RNA é significativa quando se considera o comportamento não-linear das curvas de demanda que a partir de uma sessão de treinamento a rede adquire a capacidade de aprendizagem e de responder às variáveis de entradas inéditas. Outros fatores relevantes são a necessidade de pouco conhecimento estatístico

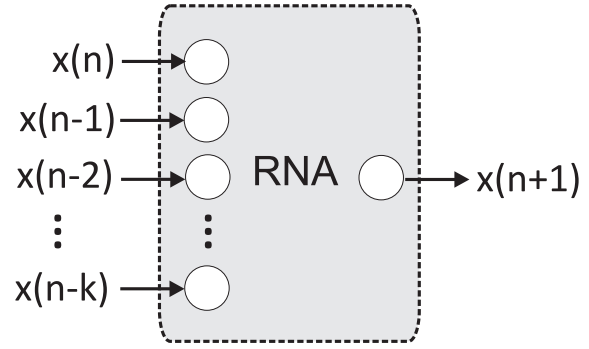


Figura 5: Modelo do Bloco RNA para previsão do próximo valor. A janela utilizada foi de 10 amostras ($k=9$).

sobre o sistema, a adaptabilidade e a generalização em relação ao ambiente no qual a rede está inserida (Castro and Castro, 2001).

Antes da aplicação do algoritmo sobre as curvas de demanda foram realizadas duas etapas de pré-processamento: tratamento dos dados e treinamento da rede.

Tratamento dos dados: foram selecionados aleatoriamente 21 dias do histórico de um ano para composição das matrizes de entrada e de alvo da Rede Neural para cada alimentador da subestação. A janela de amostras anteriores utilizada foi igual a 10 medições, ou seja, as entradas da rede são 10 medições válidas e a saída é o valor previsto, como ilustrado na Figura 5. Portanto, a matriz de entrada conta com 10 elementos e a matriz de objetivo é o valor subsequente.

Treinamento: Foi utilizado o treinamento Levenberg-Marquadt com 10 neurônios na primeira camada, 1 na camada oculta e função de ativação da tangente hiperbólica. Na etapa de treinamento foram utilizadas 80% das amostras, 10% para validação e 10% para testes.

3.3 Abordagem ARIMA

Há uma grande variedade de métodos de estimação que podem ser utilizados para a previsão de séries temporais. E como o perfil das curvas de carga é bastante similar, uma opção interessante é utilizar a técnica ARIMA (*Autoregressive Integrated Moving Average Model*) que tem sido amplamente utilizada em várias áreas para previsão. A técnica ARIMA modela a relação entre o valor atual e os valores anteriores, acrescidos de um ruído aleatório (He et al., 2012). A equação que descreve o modelo ARIMA(p, d, q) é dada por:

$$\hat{y}^d = c + \phi_1 y^d(n-1) + \dots + \phi_p y^d(n-p) + \theta_1 e(n-1) + \dots + \theta_q e(n-q) + e(n) \quad (5)$$

em que \hat{y}^d é o valor estimado, p é a ordem da parte autorregressiva, d é o grau de diferença da série temporal original, q é a ordem da média móvel,

$e(n - q) = y(n - q) - \hat{y}(n - q)$, $e(n)$ é o ruído branco e c é uma constante.

A escolha adequada dos parâmetros do modelo ARIMA(p, d, q) produzirá resíduos com distribuição de média igual a 0 e nenhuma correlação de tempo. Em aplicações práticas, a ordem do modelo é selecionada de acordo com Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). A ideia é escolher o modelo ARIMA que minimize tais critérios (Carvalho et al., 2016). A obtenção do melhor modelo foi feita de acordo com o proposto em (Hyndman and Athanasopoulos., 2013).

4 Resultados e Discussões

Os dados expostos nesta seção são de uma subestação real localizada no município de Cuité no estado da Paraíba. Os resultados serão apresentados em três cenários: no período de 1 dia, 1 ano e 6 anos. A avaliação em um dia permite conhecer melhor o comportamento do perfil da carga e do algoritmo de correção em um dos alimentadores. Os resultados de 1 ano serão expostos para todos os alimentadores. E, por fim, as técnicas de correção serão expostas a um período de 6 anos com a finalidade de avaliar o desempenho estatístico e a robustez dos métodos propostos.

4.1 Primeiro Cenário: 1 dia

Foram calculados os erros relativos de cada valor corrigido em relação ao dado original e foi calculado o Erro Médio Percentual Absoluto (MAPE – *Mean Absolute Percentual Error*), utilizado para avaliar o desempenho da estimação e dado por:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{R_t - F_t}{R_t} \right| \quad (6)$$

em que R_t é o valor real, F_t é o valor estimado e n é o número de elemento estimados.

No primeiro cenário de teste, é analisada a injeção de 5 *outliers* aleatórios não-consecutivos na curva do Alimentador NFL, ilustrada na Figura 6. A Tabela 1 indica os erros relativos médios de cada uma das técnicas desenvolvidas e para esse cenário a Abordagem ARIMA foi levemente superior. Os *outliers* ocorridos às 17h30 e às 18h30 fazem parte da região do perfil de carga em que há uma maior variação da derivada e que são zonas críticas para as técnicas de correção. A Abordagem que apresenta o pior desempenho nas regiões de maior variação é Técnica *Fuzzy* que responde inferiormente às grandes mudanças, como pode ser observada pela Figura 6 e pelo erro relativo igual a 3,90% e 3,21%, respectivamente.

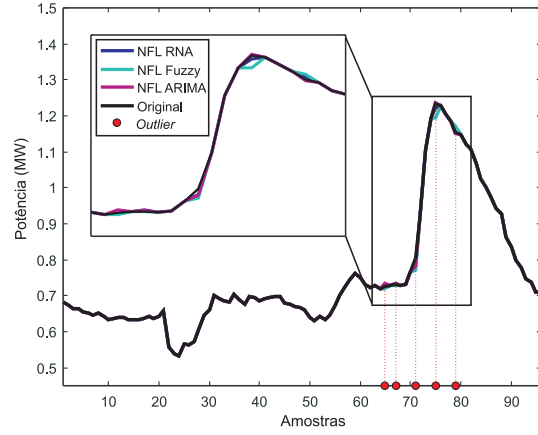


Figura 6: Primeiro cenário de teste: injeção de 5 *outliers* no intervalo de 1 dia no Alimentador NFL.

Tabela 1: Erros Relativos da injeção de 5 *Outliers* no intervalo de um dia.

Horário [Índice]	Erro R. Fuzzy (%)	Erro R. RNA (%)	Erro R. ARIMA (%)
16h00 [65]	0,65	0,43	1,51
16h30 [67]	0,05	0,76	0,52
17h30 [71]	3,90	2,51	2,61
18h30 [75]	3,21	0,89	0,35
19h30 [79]	1,48	0,88	0,33
Erro R. Médio (%)	1,86	1,09	1,06
Média MAPE (%)	0,37	0,22	0,21

Tabela 2: Fator de Ocorrência de *Outliers* e Diagnóstico do Sistema de Medição do Alimentador.

Intervalo	Diagnóstico
$f_o = 0$	Ótimo
$f_o \leq 5\%$	Aceitável
$5 < f_o \leq 10\%$	Tolerável
$10 < f_o \leq 30\%$	Inaceitável
$f_o > 30\%$	Crítico

4.2 Segundo Cenário: 1 ano

O relatório gerado pelo módulo proposto para o Sistema Supervisório indica para o Alimentador em questão: o número de *outliers* e seus horários, além dos valores corrigidos.

O algoritmo proposto, além de identificar e corrigir os *outliers*, gera um relatório com o número de falhas e um diagnóstico do sistema de medição do alimentador obtido via consulta direta à Tabela 2 em que f_o é o fator de ocorrência dado por:

$$f_o = \frac{n_o}{N} \quad (7)$$

com n_o sendo o número de ocorrência de *outliers* e N sendo o número total de amostras no intervalo de tempo estudado.

No segundo cenário de teste, foi feita a injeção aleatória de *outliers* no intervalo de um ano nas curvas dos quatro alimentadores da subestação. A Tabela 3 é similar ao relatório gerado pelo algo-

ritmo após o ciclo de um ano e contém: a Subestação e o Alimentador, o código do Alimentador referente a região de distribuição, número de falhas por ano, o fator de ocorrência e o diagnóstico do sistema de medição. Além desta avaliação, é gerado um relatório para cada alimentador com a hora e dia em que ocorreu o *outlier*. Vale ressaltar que para cada *outlier* é acionada uma *flag* para o Sistema SCADA que fica habilitada enquanto não houver dados válidos. Se o erro persistir, o Sistema Supervisório permanecerá sinalizando indicando que os dados estão sendo gerados pelo algoritmo de correção.

Para o segundo cenário de testes, o Alimentador que apresentou o pior sistema de medição foi o Alimentador 21L3-BSR com 32,11% de falhas nos dados totais do período de um ano, equivalente a 35040 amostras. E o que apresentou o melhor fator de ocorrência foi o Alimentador 21L1-BSO com apenas 3,47% das amostras com problemas.

Os erros relativos médios são indicados nas Tabela 4. Neste cenário, a técnica de correção que apresentou o menor erro relativo médio em todos os alimentadores foi a Abordagem RNA. Quando a análise é feita desconsiderando a RNA, a Abordagem *Fuzzy* foi levemente superior em 2 alimentadores, 21L1-BSO e 21L2-CTE, enquanto que a Abordagem ARIMA foi superior em 21L3-BSR e 21L4-NFL.

4.3 Terceiro Cenário: 6 anos

O terceiro cenário de testes foi realizado para um período de 6 anos em que as técnicas são analisadas quanto à robustez de responderem a uma tendência associada ao longo período de tempo.

A Tabela 5 indica o número de *outliers* presentes em cada alimentador no período de 6 anos. Apenas o Alimentador 21L1 (BSO) apresentou um fator de ocorrência aceitável. Os dados dos outros alimentadores estavam deturpados com uma quantidade considerável de dados errôneos e em um caso real, deveria ser considerada a correção ou troca do equipamento de medição.

Os erros relativos médios de cada técnica são mostrados na Tabela 6 e neste cenário, a Abordagem RNA foi superior nos Alimentadores BSO e CTE enquanto que para os Alimentadores BSR e NFL, a Abordagem *Fuzzy* apresentou o menor erro relativo médio.

As técnicas de correção utilizadas mantiveram o erro relativo médio próximo aos valores encontrados no período de 1 dia e de 1 ano. As regras da Abordagem *Fuzzy* foram construídas para o perfil de carga desta subestação então desde que ele não se altere, a técnica permanecerá respondendo positivamente. A Abordagem RNA por utilizar uma janela de amostras anteriores reduzida favorece a previsão a curtíssimo prazo, assim como a Abordagem ARIMA que tem como entrada do modelo

autorregressivo 96 medições anteriores.

Tabela 3: Relatório gerado após o período de um ano com as informações sobre cada alimentador da subestação.

Alimentador	Número de falhas	Fator de Ocorrência (%)	Diagnóstico
21L1-BSO	1244	3,47	Aceitável
21L2-CTE	9110	22,87	Inaceitável
21L3-BSR	13666	32,11	Crítico
21L4-NFL	3048	8,37	Tolerável

Tabela 4: Erros Relativos Médio para cada técnica de correção no período de um ano.

Alimentador	Erro R. Médio Fuzzy (%)	Erro R. Médio RNA (%)	Erro R. Médio ARIMA (%)
21L1-BSO	2,60	2,22	2,49
21L2-CTE	4,25	3,60	4,10
21L3-BSR	3,82	3,61	4,05
21L4-NFL	2,58	2,25	2,75

Tabela 5: Relatório gerado após o período de seis anos com as informações sobre cada alimentador da subestação.

Alimentador	Número de falhas	Fator de Ocorrência (%)	Diagnóstico
21L1-BSO	4293	2,02	Aceitável
21L2-CTE	94000	35,99	Crítico
21L3-BSR	54775	22,92	Inaceitável
21L4-NFL	48357	20,52	Inaceitável

Tabela 6: Erros Relativos Médios do Terceiro Cenário de Testes - 6 anos.

Alimentador	Erro R. Médio Fuzzy (%)	Erro R. Médio RNA (%)	Erro R. Médio ARIMA (%)
21L1-BSO	2,62	2,48	2,49
21L2-CTE	4,60	4,22	4,61
21L3-BSR	4,03	4,35	4,67
21L4-NFL	2,84	3,17	2,87

5 Conclusão

O desenvolvimento do módulo inteligente para a detecção e correção de *outliers* do Sistema Supervisório de uma subestação implica em vantagens associadas à qualidade das curvas de demanda e aumento da robustez do sistema SCADA.

A proposta de desenvolvimento do módulo complementar utilizou três técnicas de correção de *outliers* (RNA, *Fuzzy* e ARIMA), sendo a abordagem baseada em RNA a mais eficiente por conseguir acompanhar as variações da curva de demanda, principalmente, no horário de ponta. Nos estudos de casos apresentados, foi observado que as técnicas desenvolvidas apresentaram um erro relativo médio inferior a 5% para um período de análise de 6 anos.

Adicionalmente, neste trabalho foi proposto um fator de ocorrência de *outliers*, que indica a

frequência de ocorrência de *outliers* em um intervalo de tempo e a partir desta avaliação é possível determinar o diagnóstico do sistema de medição. Portanto, esta informação fornecida pelo Sistema Supervisório pode ser usada como tomada de decisões para os especialistas das concessionárias para o aprimoramento dos sistemas de automação e controle das subestações.

Referências

- Alves, W., Martins, D., Bezerra, U. and Klautau, A. (2017). A hybrid approach for big data outlier detection from electric power scada system. *IEEE Latin America Transactions*, vol. 15, no. 1, pp. 57-64.
- Anwar, A., Mahmood, A. and Pickering, M. (2016). Estimation of smart grid topology using scada measurements. *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Sydney, NSW, pp. 539-544.
- B. C. Neagu, G. Grigoras, F. S. (2017). Outliers discovery from smart meters data using a statistical based data mining approach. *10th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Bucharest, pp. 555-558.
- Bikmetova, R., Raja, M. Y. A. and Sane, T. U. (2017). Infrastructure and applications of internet of things in smart grids: A survey. *North American Power Symposium (NAPS)*, Morgantown.
- Carvalho, L. M., Teixeira, J. and Matos, M. (2016). Modeling wind power uncertainty in the long-term operational reserve adequacy assessment: A comparative analysis between the naïve and the arima forecasting models. *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Beijing, pp. 1-6.
- Castro, F. C. C. and Castro, M. C. F. (2001). *Redes Neurais Artificiais*, Rio Grande do Sul: PUCRS.
- Charytoniuk, W. and Chen, M. (2000). Very short-term load forecasting using artificial neural networks. *IEEE Trans. Power Systems*, vol. 15, pp. 263-268.
- D. C. R. Melo, A. R. G. C. (2014). Uma nova abordagem para detecção e correção de outliers em séries temporais: Aplicação em consumo de energia elétrica. *Simpósio Brasileiro de Sistemas Elétricos SBSE*.
- Georgescu, V. C. (2013). Optimized scada systems for electrical substations. *8th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, Bucharest, pp. 1-4.
- H. He, J. Y. (2016). Cyber-physical attacks and defences in the smart grid: a survey. *IET Cyber-Physical Systems: Theory and Applications*, vol. 1, no. 1, pp. 13-27.
- Han, D. and Li, X. (2012). The forecasting of electrical consumption proportion of different industries in substation based on scada and the daily load curve of load control system. *International Conference on Computer Distributed Control and Intelligent Environmental Monitoring*, Hunan, pp. 738-741.
- He, H., Liu, T., Chen, R., Xiao, Y. and Yang, J. (2012). High frequency short-term demand forecasting model for distribution power grid based on arima. *IEEE International Conference on Computer Science and Automation Engineering (CSAE)*, Zhangjiajie, pp. 293-297.
- Hyndman, R. J. and Athanasopoulos, G. (2013). *Forecasting: principles and practice*. *OTexts [Online]*, Disponível em: <<http://otexts.org/fpp/>>.
- Katic, N. A. (2016). Profitability of smart grid solutions applied in power grid. *Thermal Science*, Vol. 20, Suppl. 2, pp. 371-382.
- M. R. Müller, E. M. C. F. (2014). Clusterização de curvas de carga para o método de dias similares. *Simpósio Brasileiro de Sistemas Elétricos SBSE*.
- Medeiros, R. A. O. (2016). Previsão de demanda no médio prazo utilizando redes neurais artificiais em sistemas de distribuição de energia. *Dissertação de Mestrado, Dept. Elec. Eng., UFPB*.
- Mohammadreza, M. F., Hatami, M. and Parniani, M. (2015). Novel approaches for online modal estimation of power systems using pmus data contaminated with outliers. *Electric Power Systems Research*.
- Ramesh, S. (2015). Development of scada automated 33kv substation model as testing platform. *International Conference on Industrial Instrumentation and Control (ICIC)*, Pune, pp. 1442-1447.
- Varodayan, D. P. and Gao, G. X. (2010). Redundant metering for integrity with information-theoretic confidentiality. *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, pp. 345-349.