DOI:10.48011/asba.v2i1.1629

Segurança Cibernética em Smart Grids: Uma support Vector machine

Helton do Nascimento Alves*. Ben-Hur Matthews Moreno Montel**
**Departamento de Engenharia Elétrica, Instituto Federal do Maranhão (IFMA), São Luis, MA 65030-000, Brasil*

+55(98)988298192 helton@ifma.edu.br
***Engenharia Elétrica, Instituto Federal do Maranhão (IFMA), São Luis, MA 65030-000, Brasil +55(98)989191360*
benhurmatthews@hotmail.com

Abstract: The Smart Grid system is an important tool to manage sustainably, economic, reliable of the electric system. it uses the topology data, line parameters and state variable obtained by State Estimator to control the operation of network. recently was introduced a new class of error that the state estimator was not prepare to detect and correct, was defined as stealthy false data injection. This kind of error may occur on real time measurements, network topology and line parameters. due to diversity of injection place, to the best correction is taken, is vary important determinated the false data injection point. The principal component analysis is aiming to retain the signal variation e maximize the difference between the normal and anormaly operation points. The Monte Carlo simulation is used tho generate samples of measurements plans to the IEEE system of 14 and 57 bus. The results confirm that the innovation index used is a potential parameter to help in identification the data entry where the Gross Error occurs.

Resumo: O sistema de Smart Grid é uma importante ferramenta para gerenciar de forma sustentável, econômica e confiável do sistema elétrico. Ele utiliza dados de topologia de rede, parâmetros da linha de transmissão e variáveis de estado obtida pelo Estimador de Estado (EE) para controlar as operações da rede. Recentemente foi introduzido uma nova classe de erros em que o EE não está preparado para detectar e corrigir, foi definido como: Injeção Furtiva de Dados Falsos (IFDF). Esse tipo de erro ocorre nas medições em tempo real, topologia da rede e parâmetros de linha. Devido a diversidade de locais de injeção, para que a melhor medida seja tomada, é muito importante determinar o ponto de injeção. Nesse artigo é proposto um support vector machine para a detecção do ponto de injeção de dados falsos. A análise de componentes principais é aplicada para reter toda a variação de sinal e maximizar a diferença entre o ponto de operação normal e anormal. A simulação de Monte Carlo é utilizada para gerar amostras de planos de medições para os sistemas de 14 e 57 barras da IEEE. Os resultados comprovam que o índice de inovação como parâmetro em potencial para ajudar na identificação de local em que ocorreu o Erro Grosseiro (EG).

Keywords: Smart Grid; False data Injection; System power State; Artificial Neural Network; Support Vector Machine; Gross Error..

Palavras-chaves: Smart Grid; Injeção de Dados Falsos; Estado de energia do sistema; Rede neural Artificial; Support vector Machine; Erro Grosseiro.

1. INTRODUÇÃO

Aumentar a eficiência na produção transporte e consumo de energia, junto ao aumento da participação do consumidor no mercado de eletricidade, através da integração de fontes alternativas e econômicas, de modo a aumentar a confiabilidade e eficiência é um grande desafio do desenvolvimento de sistemas avançados de smart grid. A informação é uma importante ferramenta para tornar o sistema de Smart grid mais observável, controlável e automatizado, ela é necessária para o funcionamento de uma das principais ferramentas do Smart Grid o EE. Ele utiliza as informações de medições em tempo real, topologia do sistema e parâmetros de linha obtidos pelo sistema: Supervisory Control and Data Acquisition (SCADA). O EE usa esses dados para estimar o estado de energia do sistema, e com isso detectar erros de medição além de identificar, detectar, e corrigir ou eliminar dados incorretos.
A IFDF foi definida como uma nova classe de EG contra a estimação de estado no sistema elétrico de potência. em (Liu et al.2009) demonstra-se que com base no conhecimento da configuração do sistema é possível construir um ataque IFDF que passará despercebido por qualquer técnica existente até o momento de detecção de dados corrompidos. Esse assunto também é tratado em ((Liu et al. 2011; Kosut et al. 2011; Hug and Giampapa. 2012; Kosut et al. 2010; Teixeira et al. 2010; Dan and Sandberg. 2010; Ashok and Govindarasu. 2012). A partir disso, muitas pesquisas e trabalhos focaram em como detectar ataques de IFDF e proteger o modulo do EE (Kim and Poor. 2011; Hug and Giampapa. 2012). O ataque de IFDF pode ocorrer em qualquer uma das três entradas de dados:

- Grupo de medições em tempo real;
- Grupo de parâmetros de linha;
- Topologia da rede, que reúne dados sobre disjuntores, interruptores, e diagrama unifilar do sistema.

Quando um ataque IFDF é identificado, o algoritmo de detecção pode tomar as seguintes medidas:

- Alertar o controle de operação do sistema;
- Eliminar a medição em que o SFDI foi identificado;
- Identificar a entrada de dados do SE em que ocorreu o ataque SFDI e corrigir, a fim de evitar estimações imprecisas

As principais contramedidas tomadas pelo algoritmo são a eliminação da medição e a correção. A eliminação tem uma grande desvantagem, que é a possibilidade do EE perder a condição de observabilidade. Já na correção, é necessário identificar em qual das três entradas de dados ocorreu o ataque. Porem na literatura, esse tema não é normalmente abordado, e se considera que já se conhece o ponto de injeção, isso é um grande erro.
O problema da estimação de variáveis de estado já é conhecido, existe um método clássico conhecido como método dos mimos quadrados (MMQ), que consiste em minimizar o quadrado da diferença entre valores calculados e medidos. Porém, existem algumas situações em que ele falha, como mostrado em (Cheniae. 1996; Mili et al. 1996; Clements and Davis. 1986; Chen and Abur. 2006). Recentemente, alguns trabalhos apresentaram uma aproximação geométrica e topológica (Bretas et al. 2017; Bretas et al. 2009; Bretas et a. 2011; Bretas et al. Fevereiro 2013; N.G. Bretas e A.S. Bretas. 2015), nesses trabalhos foram propostas metodologias para compor o erro de medição e corrigi-los.

Existem diferentes métodos para a detecção de intrusão. Usaremos dois trabalhos para comparar com os resultados obtidos nesse trabalho: (bretas et al. 2017), que mostra uma abordagem inovadora em segurança cibernética, e (Lourenço et al. 2015) que é focado na topologia e processamento de dados falsos na estimação de dados generalizados, para determinar o erro, ele usa a interpretação geométrica do vetor de multiplicadores de Lagrange.

Para resolver esse problema, as principais contribuições desse trabalho são:

- Enquanto diferentes métodos têm sido usados para a detecção de intrusão, ao melhor conhecimento dos autores, alguns artigos lidam especificamente com a identificação de dados de entrado do EE onde foi lançado um ataque IFDF, sendo um tema que requer mais estudos.
- Enquanto muitas técnicas presentes na literatura usam os resíduos de medição obtidos pelo SE clássico com mínimos quadrados como parâmetro principal em sua análise, o algoritmo proposto usa o erro de medições compostas como parâmetro principal de sua análise.
- Para identificar em qual entrada de dados do EE foi lançado o ataque IFDF, é desenvolvido um algoritmo geral aplicado a qualquer smart grid baseado apenas no vetor de estado estimado obtido pelo SE.

Também é observada a aplicação da Análise de Componentes Principais (PCA) na base de dados utilizada no algoritmo de identificação e resultados relatados.

2. FORMULAÇÃO DO PROBLEMA

2.1 Injeção furtiva de dados falsos

O estimador de estado clássico, utiliza o vetor de resíduo de medição para identificar EG. Esse método depende da topologia do sistema, das medições em tempo real e dos parâmetros de linha. Caso um invasor tenha conhecimento dessa base de dados, um EG pode ser lançado da seguinte forma:

$$Z_{Ataque} = Z + a \tag{1}$$

Onde a reflete uma IFDF sobre as medições em tempo real, topologia do sistema ou parâmetros de linha. Nesse caso, o vetor de variáveis estimadas pode ser escrito da seguinte forma:

$$X_{Ataque} = \widehat{X} + c \tag{2}$$

com isso, podemos reescrever a equação do resíduo de medição:

$$R(\hat{x}) = Z - H\hat{x} + a - Hc \tag{3}$$

Observa-se que caso $a=Hc$, o vetor de resíduo de medição não muda, logo o teste de identificação de invasor falha. Para esse tipo de EG é conhecido como IFDF.

2.2 Conceito inovador

Este trabalho é baseado no conceito inovador introduzido por Bretas et al. (2009; 2013 a, b; 2015; 2017). Considerando o sistema observável, podemos decompor o espaço vetorial das medições, na soma direta de dois subespaços vetoriais,

$$R^m = R(H) \oplus R(H)^{(\cdot \perp \cdot)} \tag{4}$$

em que o range space de H é um subespaço vetorial de dimensão N que pertence a R^m e é seu complemento.
No estimador de estado clássico, o objetivo é minimizar a função de erro. Esse processo pode ser entendido como a projeção do vetor de mismatch de medição Δz em $R(H)$. Em Bretas et al. (2009; 2013 a, b; 2015; 2017) é introduzido o operador linear P que performa essa projeção da seguinte forma:

$$P = H[H^T W H]^{-1} H^T W \tag{5}$$

Tendo conhecimento de (4) e (5), a formulação linear do estimador de estado pode ser usada para dividir o vetor erro de medição e em duas componentes: uma detectável (vetor de medição residual) e uma indetectável (ortogonal a componente detectável). Elas são obtidas da seguinte forma:

$$e_D = (I - P)e \tag{6}$$

$$e_U = P_e \tag{7}$$

$$\| e_i \|_W^2 = \| e_{Di} \|_W^2 + \| e_{Ui} \|_W^2 \quad (8)$$

onde: *eD* é a componente detectável, *eu* é a componente indetectável e *ei* é o erro de medição composto.

A fim de encontrar a componente indetectável e compor o erro de medição da iº medição, usamos o índice de inovação II:

$$II_i = \frac{\| e_{Di} \|_W}{\| e_{Ui} \|_W} \quad (9)$$

Uma medição com um baixo índice de inovação indica que uma grande componente de seu erro não é refletida no seu resíduo, como obtido no estimador de estado clássico. Consequentemente, até quando suas medições possuírem EG, seus resíduos serão relativamente pequenos.

Usando (8) e (9) é possível estimar o erro de medição composto para a iº medição baseado na matriz de covariância residual. Ele é obtido da seguinte forma:

$$\| e_i \|_D^2 = \| e_{Di} \|_W^2 + \left\| \frac{e_{Di}}{II_i} \right\|_W^2 \quad (10)$$

$$\| e_i \|_W = CME_i^N = \| e_{Di} \|_W \sqrt{(1 + \frac{1}{II_i^2})} = \frac{r_i}{\sigma_i} \sqrt{(1 + \frac{1}{II_i^2})} \quad (11)$$

Atribuímos pesos para as medições, pois como descrito em N.G. Bretas e A.S. Bretas (Dezembro 2015) todas as medições podem conter erros, ou seja, os estágios de detecção não importam no quão confiável a medição é.

Os pesos são atribuídos da seguinte forma:

$$W_{ii} = \frac{1}{(0.1 z_i)} \quad (12)$$

3. ALGORITMO PROPOSTO

Nesse artigo é proposto uma técnica baseada em aprendizagem de máquina, para a detecção de ponto de injeção de dados falsos em Smart Grids, baseado no conceito inovador. O método é baseado numa support vector machine distribuída.

Planos de medições foram gerados pela simulação de Monte Carlo, eles seguem um modelo de dados, em que dividimos em três grupos, que correspondem aos três grupos de dados em que um EG pode ser lançado.

O conceito inovador que serve como base para esse trabalho apresenta dois índices (II e CMEN), que serão utilizados como parâmetros de entrada.

A análise de componentes principais é usada para reduzir a base de dados. O algoritmo não foca no sistema de comunicação, e sim em uma técnica para a detecção de EG em um sistema de transmissão controlado por smart grid, onde o sistema de comunicação informa os dados.

3.1 Modelo de dados

Nesse trabalho utiliza-se o estimador de estado clássico associado ao índice de inovação, e o erro de medição composto ja calculado. Os testes mostram a dependência de qual entrada de dados é adicionado o erro grosseiro, e mostra um padrão na média e desvio padrão do II e CMEN.

O plano de medições é formado por dois grupos de medições: O Grupo de Medições Padrão (medições presentes na subestação) (GMPa), e o Grupo de Medições Probabilísticas (GMPr) (medições definidas pela simulação de Monte Carlo). São consideradas 400 amostras de Monte Carlo por caso, elas são feitas para determinar aleatoriamente um plano de medições. o plano só é escolhido se for observável.

Todo plano de medição tem um ruído associado de ±3σ sem que seja considerado um EG. Para montar uma base de dados, foram geradas 3000 amostras de Monte Carlo com as seguintes características:

- 1000 amostras considerando múltiplos erros de medição (entrada de dados 1) com EG variando entre 4σ e 14σ. Foi considerado que o conjunto de medições com EG está entre 3 e 20 medições, o conjunto e suas magnitudes foram escolhidas aleatoriamente, sem a presença de medições críticas ou grupo crítico.
- 1000 amostras considerando erros de parâmetros com EG (entrada de dados 2) variando entre 4σ e 14σ nas impedâncias em série da linha de transmissão. O EG foi adicionado diretamente na magnitude dos parâmetros em série da linha. foi considerado que o conjunto de linhas de transmissão com EGs está entre 1 e 3 (escolhido aleatoriamente assim como sua magnitude).
- 1000 amostras considerando um erro de exclusão topológica (entrada de dados 3), onde uma linha foi excluída. A linha de transmissão excluída foi escolhida aleatoriamente.

3.2 Parâmetros de entrada

Para avaliar a II e CMEN como parâmetros de entrada para a identificação de qual entrada de dados ocorreu o EG, consideramos suas médias e desvios padrões máximos e mínimos, os dados foram obtidos com a simulação de Monte Carlo com dois planos de medições.

Na tabela 1, considera-se um plano completo com 122 medições sendo elas: 14 de injeção de potência ativa, 14 de injeção de potência reativa, 40 fluxos ativos, 40 fluxos reativos e 14 magnitudes de tensão.

Tabela 1. Valores mínimos e máximos da média e desvio padrão para 3000 execuções (122 medições)

Status	Índice	Média Min.	Média Max.	STD Min	STD Max.
EG Entrada de dados 1	CME	0.861	2.87	0.927	4.147
EG Entrada de Dados 2		1.401	23.25	1.338	19.730
EG Entrada de Dados 3		6.542	51.593	16.326	59.484
EG Entrada de Dados 1	II	2.44	2.55	1.69	1.85
EG Entrada de Dados 2		2.49	2.56	1.77	1.90
EG Entrada de Dados 3		2.43	45.57	1.71	201.36

Na tabela 2 é considerado um plano de medição com 81 medições, sendo elas: GMPa: injeção de potências ativa e reativa, e magnitude da tensão na barra 1. GMPr: 78 medições definidas pela simulação de Monte Carlo, considerando uma função de distribuição normal para injeção de potência ativa e reative, e fluxo de potência.

Tabela 2. Valores mínimos e máximos da média e desvio padrão para 3000 execuções (81 medições)

Status	Índice	Média Min	Média Max	STD Min	STD Max
EG Entada de Dados 1	CME	0.928	2.97	1.112	4.6
EG Entrada de Dados 2		1.54	26.2	1.32	18.6
EG Entrada de Dados 3		3.69	49.3	14.6	52.19
EG Entrada de Dados 1	II	2.42	2.55	1.69	1.85
EG Entrada de Dados 2		2.50	2.58	1.76	1.93
EG Entrada de Dados 3		2.43	45.57	1.67	193.4

Os resultados mostram que II e o CME tem potencial para ser usado como variáveis para a detecção de qual entrada de dados foi lançado um EG. Pois os valores de média e desvio padrão máximo e mínimo apresentaram um certo padrão nos dois planos de medições.

3.3 Detecção de qual entrada de dados ocorreu o EG

No conceito inovador introduzido por Bretas et al. 2017, foi apresentado um método analítico para a identificação da entrada de dados em que ocorreu o EG. O método é iniciado depois da detecção de um ataque cibernético, e propõe a identificação desse ataque analisando a característica de um ataque especifico, da seguinte forma:

1. Um ciberataque de parâmetro na linha i-j espalhará o erro em todas as equações em que este parâmetro está presente, assim, o respectivo fluxo de potência ativo ou reativo i-j e j-i apresentará erros com valores de alta magnitude, bem como as injeções nos barramentos limite. Este ataque é identificado analisando as primeiras CMENs maiores >3;
2. Um ciberataque de topologia de sistema pode ser considerado como um caso extremo de um ciberataque de parâmetro, também espalhando o erro na vizinhança da linha de transmissão. Neste caso, porém, considerando que ocorreu um erro topológico de exclusão, ou seja, a linha operacional i-j foi configurada como estando offline, o respectivo fluxo de potência ativa ou reativa i-j j-i não aparecerá, uma vez que a linha foi excluída. Por outro lado, as injeções de energia i e j apresentarão valores CMEN muito altos, devido ao desequilíbrio de potência gerado pela exclusão da linha de transmissão. Este ataque é identificado analisando as primeiras CME^N Maiores>10.
3. Se nenhum dos testes acima identificaram o ataque cibernético, é considerado um ataque cibernético de medição.

3.4 Análise de componentes principais

Na maioria das vezes, a base de dados possui uma grande quantidade de atributos redundantes e irrelevantes, tornando-a bem maior do que o ideal em aspectos computacionais. Para isso técnicas de redução de dados podem ser aplicadas no conjunto original, tornando-os bem menor em volume, porém mantendo sua integridade original e suas informações principais.

Dentro das técnicas de redução de dados, optamos pela análise de componentes principais, que é um procedimento de redução de variáveis que usa utiliza transformações ortogonais para converter um conjunto de observações de possíveis variáveis correlacionadas num conjunto de valores de variáveis linearmente não correlacionadas, que é chamada de componente principal. Essa transformação é definida de forma que a primeira componente principal tem a maior possibilidade de variação, e os componentes que a sucedem por sua vez, tem a maior variação possível sobe as restrições, que é ortogonal a componente que a sucede. Após essa redução de dimensionalidade, é possível obter uma reconstrução aproximada das variáveis originais de um número pequeno de componentes principais que contêm a informação principal.

Existem muitas abordagens computacionais para essa análise, as mais modernas utilizam decomposição de autovetor (DAv) para casos de matriz quadrada, e decomposição de valor singular (DVS) para casos de matriz retangular.

A forma comum do conceito de PCA quando desenvolvido através da DVS de uma matriz de dados de valor complexo é obtida da seguinte forma:

- Organizar o conjunto de dados: Definir uma única matriz X de dimensões n x p. Com cada coluna representando uma única observação agrupada das p variáveis.
- Calcular os desvios em relação à média (B): centros X subtraindo os meios de coluna.
- Encontrar a decomposição do valor singular de B. [U,S,V] = DVS(B) produz uma matriz diagonal S, da mesma dimensão que B e com elementos diagonais não negativos em ordem decrescente, e matrizes unitárias U e V de modo que B = U*S*V'.
- Definir COEFF=V como os coeficientes do componente principal, tambem conhecido como carregamentos.
- Definir SCORE=U*S como os escores principais do componente, isto é, a representação de X no espaço do componente principal. As linhas de SCORE correspondem a observações, colunas a componentes.
- Use X_{new} =SCORE(:,1:z)*COEFF(:,1:z)^T + mean(X) para obter um vetor reconstruído aproximadamente considerando apenas os primeiros componentes principais de z.

3.5 Método de identificação: Support Vector Machine

Uma support vector machine, é um conjunto de métodos de aprendizagem de máquina supervisionada usada para analisar padrões e classificação de dados (Vegapnikk 1998).

Originalmente, a support vector machine é aplicada para a classificação da classe binária, mas existem abordagens que permitem sua aplicação em múltiplas classes. A abordagem mais comum, consiste em transformar um problema de múltiplas classes em vários problemas de classe binária. As estratégias para a solução desses problemas incluem: estratégia um contra todos, que ajusta um classificador por classe, e a estratégia um contra um, que visa ajustar um classificador por par de classe.

A support vector machine é um algoritmo classificador de risco estrutural, ele se mostrou superior aos algoritmos empíricos tradicionais de minimização de risco usados pelas redes neurais artificiais convencionais, pois ele minimiza um limite superior em vez de um risco esperado, enquanto os métodos tradicionais minimizam os erros nos dados de treinamento.

A support Vector Machine busca determinar a posição de otimização de um separador linear de hiperplano entre as classes de dados binarias. O ponto de dados mais próximo a esse hiperplano ótimo é conhecido com support vector.

Existem situações em que essa separação não é tão simples. Nesses casos, a support vector machine pode suavizar a margem, significando num hiperplano que separa muitos, mas não todos os pontos de dados. Esse processo envolve a adição de uma variável vetorial de folga não negativa ξ e um parâmetro de penalidade ajustável C. ξ é o limite superior no número de erros de treinamento e C é o parâmetro que controla a troca entre margem e erro de treinamento. Um alto valor de C indica que há mais importância na classificação correta de todos os dados de treinamento, já um valor baixo, resulta numa maior flexibilidade do hiperplano, que tenta minimizar a margem de erro.

Muitos problemas em aplicações reais possuem uma margem entre categorias não linear e mal separado por um hiperplano de separação linear, mesmo com o uso de margens flexíveis. Para tentar resolver esse problema, a support vector machine aplica funções de núcleo para mapear o conjunto de dados em um espaço de recursos dimensionais mais avançado, buscando torná-lo linear nesse novo espaço. As funções de núcleo mais utilizadas em support vector machines são: linear, polinomial, função de base radial e sigmoide. Em (Burges. 1998) apresenta-se informações introdutórias sobre support vector machines.

A abordagem da support vector machine de multiclasse usada nesse trabalho baseia-se em:

- Design de codificação um contra um, nesse caso, são definidas $k(k-1)/2$ support vector machine, onde k é o número de rótulos de classe exclusivas.
- Modelo de Código de Saída com Correção de Erro (CSCE), que reduz o problema de classificação com três ou mais classes de classificadores binário. Para isso, é definida uma matriz de codificação $M \in \{-1, 0, 1\}$, em que o valor zero indica que a classe não foi classificada para o treinamento de um classificador particular, isso cria diferentes limites de decisão, resultando numa maior precisão na classificação de problemas multiclasse.
- Esquema de decodificação da função de perda binaria. Determina como as previsões dos classificadores binários são agregados, ou seja, uma nova observação é atribuída a classe c, que minimiza a agregação das perdas para todos as $k(k-1)/2$ support vector machine binarias. Nesse trabalho, foi usada a função de perda de Hinge.

4. APLICAÇÃO

Os testes numéricos foram realizados utilizando os sistemas de 14 e 57 barras da IEEE. As médias agregadas da simulação de Monte Carlo obtidos pelo algoritmo da support vector machine são apresentados nas tabelas 3 e 4.

Erros aleatórios são adicionados a todas as medições. 3000 amostras são geradas e usadas como conjunto de treinamento (1000 amostras com EG em cada grupo). 150000 amostras são geradas para testar a rede treinada (50000 amostras com EG em cada grupo de medição). E são considerados até 10 medidores fora de serviço no sistema IEEE-14 barras, e 39 no IEEE-57.

Tabela 3. Identificação da entrada de dados com EG (%) IEEE-14 bus

Grupo de Med.	Dados de entrada aplicando PCA			Entrada de Dados original		
	EG entr. dados 1	EG entr. dados 2	EG entr. dados 3	EG entr. dados 1	EG ientr. dados 2	EG entr. dados 3
95	100.00	99.75	100.00	99.89	98.3	99.56
94	100.00	99.69	100.00	99.87	98.5	99.4
93	100.00	99.73	100.00	99.73	98.22	99.23
92	100.00	99.71	100.00	99.63	97.9	99.15
91	100.00	99.67	100.00	99.52	97.94	99.07
90	100.00	99.65	100.00	99.43	97.99	98.96
89	100.00	99.62	100.00	99.23	98.14	99.37
88	100.00	99.67	100.00	99.03	98.02	99.42
87	100.00	99.72	100.00	98.75	97.75	99.22
86	100.00	99.79	100.00	98.89	98.1	99.27
85	100.00	99.67	100.00	97.5	97.86	99.21

O desvio-padrão máximo obtido para o sistema IEEE-14 barras é de 0,73% como mostra a Tabela 3. A rede é treinada considerando planos de medição com 95 medições (GMPa: injeção de energia ativa e reativa e magnitude de tensão no barramento 1; GMPr: 92 medições definidas pela simulação de Mote Carlo, considerando uma função de distribuição normal de injeção de potência ativa e reativa e fluxo de energia).

A tabela 4 mostra os resultados obtidos para o sistema IEEE-57 barras a rede foi treinada considerando planos de medição com 339 medições (GMPa: injeção de energia ativa e reativa e magnitude de tensão em ônibus 1, 2 e 3; GMPr: 330 medições definidas pela Simulação de Monte Carlo, considerando uma função de distribuição normal de injeção de potência ativa e reativa e fluxo de potência e magnitude de tensão). O desvio padrão máximo obtido foi de 0,65%.

Tabela 4. Identificação da entrada de dados com EG (%) IEEE-57 bus

Grupo de Med.	Dados de entrada aplicando PCA			Entrada de Dados original		
	EG entr. dados 1	EG entr. dados 2	EG entr. dados 3	EG entr. dados 1	EG ientr. dados 2	EG entr. da-dos 3
339	100.00	99.85	100.00	100.00	98.5	97.26
335	100.00	99.59	100.00	100.00	97.9	97.24
331	100.00	99.76	100.00	100.00	98.22	97.21
327	100.00	99.84	100.00	100.00	98.1	97.19
324	100.00	99.67	100.00	100.00	97.94	97.07
320	100.00	99.65	100.00	99.63	98.33	97.16
316	100.00	99.82	100.00	99.72	98.94	97.17
312	100.00	99.77	100.00	99.03	98.62	97.12
308	100.00	99.72	100.00	99.45	98.75	97.23
304	100.00	99.83	100.00	99.89	98.1	97.17
300	100.00	97.87	100.00	99.5	97.00	97.21

Para melhorar o controle e funcionamento do sistema, pequenas reconfigurações na topologia da linha são consideradas (transferência de um ramo entre nós ou isolamento de um ramo pelo seccionalizador). Entretanto, o algoritmo não é treinado para isso, e considera-se apenas a topologia do sistema original. Os resultados são apresentados nas tabelas 5 e 6.

Tabela 5. Identificação da entrada de dados com EG (%) IEEE-14 bus

Pequenas Mudanças	Dados de entrada aplicando PCA			Entrada de Dados original		
	EG entr. dados 1	EG entr. dados 2	EG entr. dados 3	EG entr. dados 1	EG ientr. dados 2	EG entr. dados 3
Nenhuma	100.00	99.34	99.95	99.86	98.42	99.176
LTransf.. 13-14→11-14	100.00	99.45	99.44	99.33	98.65	98.97
Eliminação Linha 10-14	100.00	99.51	99.21	96.32	95.42	95.7
L. Transf. 4–5 → 3-5	100.00	99.47	99.1	98.3	99.2	98.5
Eliminação Linha 2-5	100.00	99.32	99.28	99.95	99.1	98.83

Tabela 6. Identificação da entrada de dados com EG (%) IEEE-57 bus

Pequenas Mudanças	Dados de entrada aplicando PCA			Entrada de Dados original		
	EG entr. dados 1	EG entr. dados 2	EG entr. dados 3	EG entr. dados 1	EG ientr. dados 2	EG entr. dados 3
Nenhuma	100.00	99.342	99.948	97.56	98.29	99.1
L. Transf. 29–52 → 26-52	100.00	99.25	99.14	98.77	98.32	99.03
Eliminação Linha 22-38	100.00	99.29	99.31	97.89	98.46	98.9
L. Transf. 29–52 → 26-52 11-13 →11-10	100.00	93.35	86.1	100.00	62.7	94.7
Eliminação Linhas 22-38 e 1-15	100.00	99.57	99.33	98.7	98.9	98.86

Os resultados mostram que o algoritmo possui uma grande generalização e uma ótima capacidade de modelagem não linear. Além disso, os resultados mostram que o erro de medição composto e o índice de inovação refletem melhor as características do EG do que o índice de resíduo normalizado. A tabela 7 compara os dados.

Tabela 7. Resultados para parâmetros de entrada baseados em índice residual normalizado e índice de inovação.

Parâmetros de entrada baseado em	Entrada de dados baseada no PCA	Entrada de dados original
	Entrada de dados identificada corretamente	Entrada de dados identificada corretamente
Índice residual normalizado	93.7 %	92.5%
Índice inovador	99..6 %	99.1 %

A tabela 8 compara os resultados obtidos pelo algoritmo proposto e pelo conceito inovador introduzido por Bretas et al. 2017, para cenários de medição gerados pela simulação de Monte Carlo.

Tabela 8. Resultados para parâmetros de entrada baseados em índice residual normalizado e índice de inovação.

Parâmetros de entrada baseado em	Entrada de dados baseada no PCA	Entrada de dados original
	Entrada de dados identificada corretamente	Entrada de dados identificada corretamente
Método analítico	22.3 %	19.9 %
Algoritmo proposto	99.7 %	99.1 %

5. CONCLUSÃO

O método proposto é baseado numa abordagem de support vector machine, e utiliza como parâmetros de entrada a média e desvio padrão obtidos pelo índice de inovação, como objetivo de recuperar os erros mascarados no processo de estimação de estado, e é um parâmetro melhor para avaliar EG.

Para avaliar o desempenho do método, foram feitos testes com os sistemas IEEE-14 e 57barras, os resultados observados chegaram próximo a 100% de identificação correta do ponto de injeção de dados falsos. As características decisivas para a obtenção desse resultado são: seu parâmetro de personalidade ajustável (evita o problema de excesso de ajuste e garante uma saída global mínima), e o uso de classificador de margem macia (confere imunidade ao ruído).

As metodologias tradicionais, baseadas em métodos analíticos possuem taxa de identificação próxima a 20%, podemos utilizar outras abordagens tendo como base os parâmetros de entrada proposto nesse trabalho, como um multilayer perceptron, os dois métodos foram testados,

porém o support vector machine obteve resultados mais satisfatórios para ambos os alimentadores e casos testados.

6. REFERÊNCIA

Ashok A., & Govindarasu M., (july 2012). Cyber attacks on power system state estimation through topology errors. *Power and Energy Society General Meeting, 2012 IEEE, pp. 1 –8.*

Bretas A.S., Bretas N.G., Carvalho B., Baeyens E., & Khargonekar P. P. (August 2017). Smart grids cyber-physical security as a malicious data attack: An innovation approach. *IEEE Trans. Power Syst. 149 210-219.*

Bretas NG., London Jr. JBA., Alberto LFC., & Benedito RAS. (July 2009). Geometrical approaches on masked gross errors for power system state estimation. *PESGM09. Calgary (USA).*

Bretas N.G., Bretas A.S., & Piereti S.A., (June 2011) Innovation concept for measurement gross error detection and identification in power system state estimation*, IET Gener. Transm. Distrib. 5 603–608.*

Bretas N.G., Piereti S.A., Bretas A.S., & Martins A.C., (February 2013). A geometrical view for multiple gross errors detection, identification, and correction in power system state estimation, *IEEE Trans. Power Syst. 28, 2128–2135.*

Bretas N.G., Bretas A.S., & Martins A.C.P., (2013). Convergence property of the measurement gross error correction in power system state estimation, using geometrical background. IEEE Trans Power Syst, 28:3729–36.

Bretas N.G., & Bretas A.S., (December 2015). A two steps procedure in state estimation gross error detection, identification, and correction, Int. *J. Electr. Power Energy Syst. 73, 484–490.*

Cheniae M.G., Mili L., & Rousseeuw PJ. (1996). Identification of multiple interacting bad data via power system decomposition*. IEEE Trans Power Syst ;11(3):1555–63.*

Chen J., & Abur A., (2006). Placement of PMUs to enable bad data detection in State estimation. *IEEE Trans Power Syst, 21(4):1608–15.*

Clements KA., & Davis PW. (1986). Multiple bad data detectability and identifiably: a geometric approach. *IEEE Trans Power Deliv, 1(3):355–60.*

Dan G., & Sandberg H. (2010). Stealth attacks and protection schemes for state estimators in power systems. *In Smart Grid Communications (SmartGridComm), First IEEE International Conference on, Oct 2010, pp. 214–219.*

Hug G., & Giampapa J.A. (2012) Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks. *IEEE Trans. Smart Grid 3 (September (3)) 1362–1370.*

Kim T., & Poor H. (2011) Strategic protection against data injection attacks on power grids. *IEEE Trans. Smart Grid 2 (2) 326–333.*

Kosut O., Jia L., Thomas R., & Tong L. (oct. 2010) Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. *In Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on, pp. 220 –225.*

Kosut O., Jia L., Thomas R., & Tong L., (Dec 2011) Malicious data attacks on the smart grid. *Smart Grid, IEEE Transactions on, vol. 2, no. 4, pp. 645–658.*

Liu Y., Ning P., & Reiter M. K., (2009) False data injection attacks against state estimation in electric power grids. *In Proceedings of the 16th ACM conference on Computer and communications security, ser. CCS '09. New York, NY, USA: ACM, pp. 21–32. [Online]. Available: http://doi.acm.org/10.1145/1653662.1653666*

Liu Y., Ning P., & Reiter M. K., (Jun. 2011). False data injection attacks against state estimation in electric power grids. *ACM Trans. Inf. Syst. Secur., vol. 14, no. 1, pp. 13:1–13:33. [Online]. Available: http://doi.acm.org/10.1145/1952982.1952995*

Lourenco E., Coelho E. & Pal B., (Nov 2015). Topology error and bad data processing in generalized state estimation, Power Systems. *IEEE Transactions on, vol. 30, no. 6, pp. 3190–3200.*

Mili L, Cheniae M, Vichare N., & Rousseau P., (1996). Robust state estimation based on projection statistics. *IEEE Trans Power Syst ,11(2):1118–27.*

Teixeira A., Amin S., Sandberg H., Johansson K., & Sastry S. (dec. 2010). Cyber security analysis of state estimators in electric power systems. *In Decision and Control (CDC), 2010 49th IEEE Conference on, pp. 5991 –5998.*

Vapnik V., (1998). Statistical Learning Theory. *John Wiley & Sons.*

Data for the test system are at: <www.ee.washington.edu/research/pstca/>.