

# O Uso de *AutoEncoders* como Fomentador de Medidas em Sistemas de Distribuição

Luiz Phillip Q. da Silva<sup>1</sup>, Julio Cesar S. de Souza<sup>1</sup>, Milton Brown Do C. Filho<sup>1</sup>

<sup>1</sup>Instituto de Computação – Universidade Federal Fluminense (UFF)  
Niterói – RJ – Brazil

**Abstract.** *The success of the automation and control functions envisaged for smart distribution grids depends on a reliable real-time distribution system monitoring. This task is performed by the distribution system state estimator, responsible for processing a set of measurements obtained through the data acquisition system. The smart grid advanced metering infrastructure (AMI) can collect voltage and power measurements readings in a regular basis, which may complement the few SCADA measurements usually available in distribution networks and benefit the state estimation process. However, due to bottlenecks in the communications infrastructure, the refresh rates of SCADA and AMI measurements are not the same. This work presents a methodology that employs an AutoEncoder to generate AMI pseudomeasurements to complement SCADA measurements when only the latter are available. Simulations carried out with a 34-bus distribution system illustrate the proposed methodology and the obtained results confirm its potential to generate pseudomeasurements.*

**Resumo.** *O sucesso de funções de automação e controle vislumbradas para as redes de distribuição inteligentes depende da supervisão confiável da rede em tempo real. Esta tarefa é realizada pelo estimador de estado da distribuição, responsável por processar um conjunto de medidas recebidas através do sistema de aquisição de dados. Em redes inteligentes, a infraestrutura avançada de medição (AMI) permite a leitura regular de medidas de tensão e potência dos consumidores, as quais podem complementar as poucas medidas usualmente presentes na rede de distribuição (medidas SCADA) e beneficiar o processo de estimação de estado. Entretanto, devido a limitações na infraestrutura de comunicação, medidas SCADA e AMI possuem taxas de amostragem e aquisição diferentes. Este trabalho apresenta uma metodologia que emprega uma rede neural do tipo AutoEncoder para a geração de pseudomedidas AMI para complementar medidas SCADA em instantes de tempo em que apenas estas últimas estão disponíveis para processamento. Simulações realizadas com um sistema de distribuição de 34 barras ilustram a metodologia proposta e os resultados obtidos confirmam o seu potencial de aplicação para a geração de pseudomedidas.*

## 1. Introdução

A supervisão da operação de redes de distribuição de energia elétrica tende a se tornar um requisito cada vez mais importante, sobretudo em tempo real por meio da função estimação de estado. Como exposto por [Lefebvre et al. 2014], essa ferramenta é essencial para o sucesso das modernas funções de automação e controle, tais como o controle

do perfil de tensão da rede, reconfiguração ótima de alimentadores, detecção de ilhamento em redes com geração distribuída, dentre outras. Entretanto, para que tal supervisão possa ser realizada eficazmente diversos desafios necessitam ser superados, muitos deles diferentes dos encontrados em sistemas de transmissão, devido a características como:

- poucas medidas são em geral coletadas para processamento e a observabilidade da rede não é alcançada sem o emprego de pseudomedidas [Mestav et al. 2019];
- a presença de desequilíbrios na rede e nas cargas dos sistemas de distribuição;
- relações entre resistências e reatâncias muito distintas das encontradas nos sistemas de transmissão, principalmente no caso das redes subterrâneas de distribuição; e
- limitações do sistema de comunicação para a transmissão de medidas que podem ser fornecidas por medidores inteligentes.

O paradigma de redes de distribuição inteligentes tem alavancado avanços nos sistemas de medição e na infraestrutura de comunicação, permitindo a aquisição e armazenamento de um grande volume de medidas, de diversas grandezas elétricas. A presença de relés digitais, unidades de medição fasoriais (PMUs — Phasor Measurement Units), dispositivos eletrônicos inteligentes (IEDs — Intelligent Electronic Devices), medidores inteligentes (SMs — Smart Meters), entre outros, representa uma oportunidade para a melhoria da observabilidade da rede. Através da Infraestrutura Avançada de Medição (AMI — Advanced Metering Infrastructure), é possível disponibilizar para processamento grandezas medidas que melhorarão a qualidade das estimativas do estado de operação da rede. Entretanto, por mais esforços que se tenham dispostos para o desenvolvimento de uma infraestrutura de comunicação adequada, a capacidade de transmissão de dados continua a ser um problema. Isso faz com que o sistema de comunicação represente um gargalo para a transmissão de dados [Li et al. 2011] e o sucesso das funções avançadas de análise, automação e controle, particularmente as relacionadas à operação em tempo real, como a estimação de estado. Isto faz com que a quantidade de medidas disponíveis num COD (Centro de Operação da Distribuição) ainda seja insuficiente para uma completa observabilidade e para a supervisão confiável da rede no ambiente de tempo real. Neste caso, o emprego de pseudomedidas se apresenta como uma alternativa para a complementação do conjunto de dados a serem processados.

Devido ao problema da não observabilidade dos sistemas de distribuição algumas técnicas têm sido analisadas de modo a mitigar os efeitos causados por esse problema. Em [Dehghanpour et al. 2019] os autores apresentam como modelos de máquinas vetoriais de relevância (RVM) podem ser utilizadas, em conjunto com a teoria dos jogos, para gerar tanto pseudomedidas quanto suas incertezas. Alguns outros autores preferem utilizar as medidas existentes para obter o estado diretamente. Os autores de [Dahale et al. 2020] mostram como o fluxo de potência linear e programação linear podem ser combinados para a obtenção do estado da rede. Em outra linha, [Mestav et al. 2019] expõe a utilização de Simulações de Monte Carlo e modelos de redes neurais profundas para obter um melhor resultado em relação à detecção de erros de medição e a estimação de estado na totalidade.

Este trabalho apresenta uma metodologia para a geração de pseudomedidas para a complementação de medidas SCADA em instantes de tempo em que apenas estas estão disponíveis para processamento. As pseudomedidas são obtidas através de uma rede neural do tipo *AutoEncoder*, a partir de uma base de dados formada por medidas SCADA

e AMI, quando ambas estão disponíveis para processamento (o que é assumido ocorrer em uma base horária). As pseudomedidas geradas permitem aumentar a frequência de instantes em que um conjunto redundante de dados fica disponível para processamento, garantindo a observabilidade e supervisão da rede. Simulações realizadas com um sistema de distribuição de 34 barras ilustram a metodologia proposta e os resultados obtidos mostram o seu potencial para a geração de pseudomedidas de boa qualidade para processamento pela função estimação de estado.

## 2. Codificadores Automáticos - (*AutoEncoders*)

AutoEncoders (AE) é uma classe de modelos de aprendizado profundo não supervisionado composto de duas estruturas, codificador e decodificador [Liu et al. 2017]. A estrutura de codificação consegue extrair características relevantes dos dados de entradas [Liu et al. 2017, Hinton and Zemel 1994]. O decodificador é responsável por, a partir das características extraídas, reconstruir os dados de entrada [Liu et al. 2017]. Essa categoria de modelo é um dos mais populares modelos generalistas em que ambas as estruturas (codificador e decodificador) são formadas por redes neurais [Jabbar et al. 2020].

Os modelos formados por redes neurais, em sua estrutura básica, são capazes de se adaptar ao conjunto de dados por uma função de custo. O objetivo dessa função é fornecer uma métrica do quão longe, ou próximo, o modelo está da reprodução fidedigna ao conjunto de dados. No caso dos AE's, a função de custo mais utilizada é a média das distâncias euclidianas entre o resultado esperado e a reconstrução efetuada pela modelo [Jabbar et al. 2020]. A Equação (1) apresenta a função de custo:

$$L_{AE} = \frac{1}{n} \sum_{i=0}^n ||x - f_{\theta}(g_{\theta}(x))|| \quad (1)$$

onde  $f_{\theta}$  representa a função do decodificador e  $g_{\theta}$  a função do codificador.

Cada função ( $f(\theta)$  e  $g(\theta)$ ) representa uma estrutura do modelo. Sendo assim suas quantidades de camadas e neurônios dependem do problema e deve ser definida pelo usuário. Após a aplicação de  $g_{\theta}$  nos dados de entrada obtemos uma nova representação dos dados de entrada ( $Z$ ). Essa novo conjunto de dados é conhecido como o espaço latente resultante da aplicação da função  $g_{\theta}$ .

Os AE's podem ser utilizados em diversas categorias de problemas. Alguns autores optam por sua utilização em detecção de anomalias [Zhang et al. 2022, Ko et al. 2021], outros preferem trabalhos na área de visão computacional para recuperação de informação e retirada de ruídos nas imagens [Yilmaz et al. 2022, Xu et al. 2022, He et al. 2022, Vijayalakshmi and Shanthakumar 2019]. Neste último caso o modelo recebe um conjunto de entrada em que há falhas ou falta de informação nas imagens e o objetivo é recriar/retirar ruído das fotos. A reconstrução pode ser ampliada para outros campos. Uma forma de se utilizar é apresentada em [Miranda et al. 2011, Krstulovic et al. 2013] onde os autores mostram como esse modelo pode ser utilizado para reconstrução de medidas quando há pouca perda de informação, cerca de 20% de perda, e também para reconstrução da topologia do sistema elétrico.

## 3. Metodologia Proposta

A infraestrutura avançada de medição em redes inteligentes permite a leitura regular de medidas de tensão e potência dos consumidores (medidas AMI), as quais, se adequa-

damente utilizadas, podem beneficiar a observabilidade da rede de distribuição e o processo de estimação de estado na totalidade. Medidas provenientes de medidores inteligentes, no nível de baixa tensão, são em geral registradas a cada 15 minutos ou em intervalos de tempo maiores. Por outro lado, o intervalo de tempo para aquisição de dados de medidas pelo sistema SCADA é de poucos segundos. Embora medidas AMI possam ser abundantes, o seu processamento em conjunto com medidas SCADA requer que uma estratégia adequada para a integração de tais dados de entrada seja utilizada [Huang et al. 2015, Zhao et al. 2020]. Isto se deve ao fato de que por longos intervalos de tempo a redundância de medidas não será suficiente para garantir a observabilidade da rede e permitir a sua supervisão. Devido às limitações no sistema de comunicação, é possível ainda que medidas AMI estejam disponíveis apenas em uma base horária.

Na metodologia proposta neste trabalho as medidas SCADA disponíveis no sistema de distribuição são complementadas por pseudomedidas que garantirão a observabilidade da rede enquanto as medidas AMI não estiverem presentes. Com base nos instantes de tempo em que as medidas AMI e SCADA estão disponíveis uma base de dados é construída, servindo como base de treinamento para uma rede neural do tipo AutoEncoder, descrito na Seção 2. Durante a fase de treinamento, o modelo empregado captura as relações entre as medidas SCADA e AMI, para conseguir reconstruir medidas AMI nos instantes de tempo em que estas não estão disponíveis no COD, a partir do conjunto de medidas SCADA observadas. Assim, é possível aumentar a frequência em que um conjunto redundante de dados estará disponível para processamento pela função estimação de estado.

Este método se divide, portanto, em duas etapas, uma etapa de construção do modelo (off-line), com maior custo computacional, e outra etapa correspondente à obtenção das pseudomedidas (on-line), que pode ser realizada no ambiente de tempo real. As subseções a seguir apresentam cada uma dessas etapas.

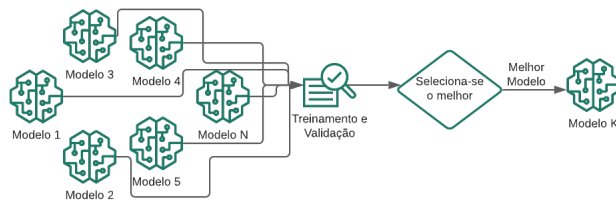
### **3.1. Construção do modelo (Off-line)**

Ao se construir um modelo existem diversos parâmetros que podem ser modificados para se obter o melhor ajuste possível aos dados disponíveis. Além disso, dependendo do modelo, há uma série de hiperparâmetros que devem ser ajustados pelo usuário, afetando todo o processo de aprendizado. Este fato faz com que um ajuste cuidadoso deva ser realizado nos hiperparâmetros visando extrair a máxima capacidade dos modelos.

A etapa off-line consiste, portanto, da seleção dos melhores hiperparâmetros capazes de obter a melhor *performante* do modelo de AE. A Figura (1) apresenta o funcionamento dessa etapa. Inicialmente é necessário definir a quantidade de testes a serem realizados, tal definição podendo ser feita por limitação de iterações ou por conjunto de hiperparâmetros previamente definidos. Após essa definição os modelos são submetidos às bases de treinamento e validação. Com uma métrica bem definida, seleciona-se o modelo que obteve a melhor métrica na etapa anterior na base de dados de validação. Na perspectiva deste trabalho, o melhor modelo será aquele a ser implementado no COD de modo a reconstruir as medidas AMI faltantes.

### **3.2. Geração de pseudomedidas (On-line)**

A fase de treinamento trata da determinação do modelo que trará mais benefícios para a supervisão do sistema de distribuição, por meio da geração de pseudomedidas. Na etapa



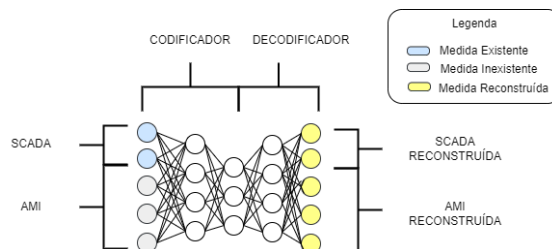
**Figura 1. Etapa Off-Line do Método de Pesquisa**

descrita nesta seção, o modelo selecionado é utilizado em tempo real para a reconstrução das medidas AMI. A Figura (2) ilustra a utilização do modelo em tempo real. Nos instantes destacados o sistema é considerado observável tendo em vista que tanto as medidas SCADA quanto AMI estão disponíveis para processamento. Entretanto, nos demais instantes o COD recebe somente as medidas SCADA, não sendo em princípio possível executar o estimador de estado. Dada essa característica o modelo de AE selecionado deverá reconstruir as medidas AMI de modo a tornar o sistema observável novamente. Com isso, o estimador de estado e outras funções importantes da operação que dependem da observabilidade podem ser executados.



**Figura 2. Etapa On-Line do Método de Pesquisa**

Para ilustrar o processo de recuperação de medidas, a Figura (3) apresenta o funcionamento do modelo quando não há medidas AMI disponíveis. O modelo recebe como entrada as medidas SCADA e nas posições correspondentes a cada medida AMI inexistente naquele momento um valor igual a zero. A partir deste conjunto de dados de entrada o modelo realiza então o processo de codificação e, após esta etapa, a decodificação. Como resultado da última etapa tem-se dentre as variáveis de saída as medidas AMI reconstruídas, garantindo um conjunto de dados que possibilita a observabilidade do sistema.

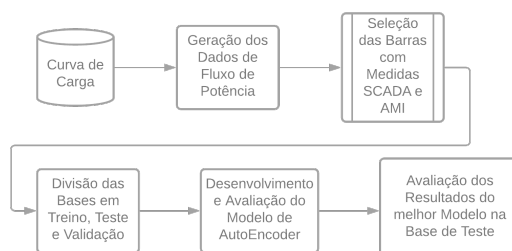


**Figura 3. Funcionamento do Modelo de AutoEncoder na Etapa On-line**

#### 4. Simulações e Resultados

O método desenvolvido neste artigo se baseia em um modelo construído com base em um conjunto de medidas provenientes do sistema de distribuição. Após a construção e

seleção do melhor modelo, este é testado e sua *performance* é avaliada empregando-se dados inéditos. Neste trabalho, testes foram realizadas com uma rede de distribuição de 34 barras apresentada em [Soares et al. 2019], para a qual dados para treinamento e validação foram gerados. Tais dados correspondem a grandezas medidas, obtidas através de um programa de fluxo de potência e posterior adição de uma perturbação aleatória. A Figura (4) apresenta o fluxograma utilizado para geração dos dados, desenvolvimento e validação dos modelos. As subseções a seguir expõem como cada uma dessas etapas foi desenvolvida.



**Figura 4. Fluxograma da Metodologia Utilizada**

#### 4.1. Geração das Bases de Dados

Os dados empregados nos testes foram gerados a partir de simulações empregando-se um fluxo de potência para redes de distribuição trifásicas desequilibradas, tendo sido considerados desequilíbrios tanto na topologia da rede quanto nas cargas. Além disso, para se gerar uma quantidade de dados suficientes para o aprendizado do modelo uma premissa sobre as cargas foi assumida. Foi estipulado que tanto o fator de potência quanto a participação das barras no consumo do sistema podem variar em  $\pm 5\%$  do valor original. Isso pode entregar uma diversidade maior aos resultados obtidos, tendo em vista que regiões que não seriam exploradas pelo modelo podem ser observadas tanto no treinamento quanto no teste.

Após a execução do fluxo de potência foi necessário transformar os dados fidedignos gerados por esse método em medidas. Para tal tarefa assumiu-se que as medidas são variáveis aleatórias, normalmente distribuídas, com média igual ao valor obtido no fluxo e desvio padrão representado como  $\sigma = \frac{z_{fluxo}}{100}$  para medidas de potência e corrente ou  $\sigma = \frac{z_{fluxo}}{300}$  para medidas de tensão. Vale ressaltar que  $z_{fluxo}$  é o valor obtido ao executar o fluxo de potência.

Foi considerado que em 10 barras da rede estão presentes medidas SCADA, estas correspondendo a medidas de fluxo de potência e reativa. As barras que contêm medidas SCADA foram selecionadas como aquelas em final de ramal ou barras que não contêm cargas. Vale ressaltar que não é possível garantir a observabilidade do sistema apenas com tais medidas SCADA, sendo necessário a geração de dados adicionais. É assumido que nas demais barras, sem medidas SCADA, existe disponibilidade de medidas AMI, correspondendo a medidas de tensão, injeção de potência ativa e injeção de potência reativa. Conforme mencionado anteriormente, as medidas AMI não estão disponíveis para processamento com a mesma frequência que as medidas SCADA e quando indisponíveis devem ser geradas utilizando-se a metodologia proposta.

Uma vez gerada uma base de dados com as medidas SCADA e AMI disponíveis para diferentes instantes de tempo, o próximo passo é a divisão da base de dados em treino, teste e validação. Nesse caso, as bases de treino e validação são utilizadas para o treinamento e seleção do melhor modelo. Enquanto a base de teste é utilizada para verificar a *performance* geral do modelo selecionado pelas bases anteriores. Vale destacar que a separação das bases foi realizada com 64% dos dados gerados para o treinamento, 16% dos dados gerados para a validação e 20% para o teste final. Além disso, o modelo receberá a todo momento 94 medidas SCADA e será responsável por reconstruir 165 medidas AMI. Isso representa a recuperação de mais de 63% das medidas.

## 4.2. Etapa Off-Line

Nesta etapa foi realizada a seleção da estrutura do modelo, tendo sido realizados vários testes variando-se a quantidade de camadas ocultas, a quantidade de neurônios em cada camada, a taxa de esquecimento para cada camada oculta da rede e a taxa de aprendizado. O objetivo foi testar diversas combinações de hiperparâmetros e verificar o modelo com menor erro de reconstrução nos dados de validação.

A seleção da melhor arquitetura para os modelos pode ser conduzida de diversas maneiras. Como os modelos de redes neurais consomem um maior tempo computacional no seu treinamento comparado a outros modelos, torna-se inviável testar todas as arquiteturas possíveis para os AE. Por esta razão, o *framework* apresentado por [Akiba et al. 2019] foi utilizado para a seleção da melhor arquitetura dentre um conjunto previamente selecionado.

**Tabela 1. Hiperparâmetros e Região Selecionada para Otimizar**

Hiper-parâmetro	Valor Mínimo	Valor Máximo
Quantidade de Camadas	3	12
Quantidade de Neurônios	15	240
Taxa de Esquecimento	0.01	0.5
Taxa de Aprendizado	$10^{-5}$	$10^{-1}$

Na biblioteca apresentada em [Akiba et al. 2019] foi selecionado o método TPE para otimização [Bergstra et al. 2011]. Por ser um método Bayesiano, nem todos os valores são testados, sendo selecionada a região que possui uma contribuição mais significativa. Vale ressaltar que todos os hiperparâmetros foram testados como distribuições log-uniformes. Além disso, a função de ativação nas camadas ocultas foi mantida como GELU. Já a função de ativação da última camada foi a Tangente Hiperbólica.

A seleção do melhor modelo baseou-se na minimização do erro percentual médio absoluto da reconstrução. Essa função de custo será apresentada a seguir.

## 4.3. Indicador de Desempenho - MAPE

O erro percentual médio absoluto (MAPE) é uma medida de erro adimensional. Essa medida possui uma característica muito interessante para engenharia, tendo em vista que em muitos casos valores percentuais são utilizados como referências. A medida de erro utilizada neste trabalho é apresentada na Equação (2).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{x_i^{real} - x_i^{reconstruído}}{x_i^{real}} \quad (2)$$

onde  $x_i^{real}$  é o valor da medida  $i$  e  $x_i^{reconstruído}$  é o valor reconstruído.

No caso das medidas, pelas premissas assumidas pode-se considerar que um MAPE menor que 3%, para medidas de potência, ou menor que 1%, para medidas de tensão, podem ser considerados satisfatórios, haja vista que esses valores representam  $\pm 3\sigma$  de tais variáveis aleatórias.

O MAPE foi utilizado nas duas etapas do método apresentado. Na etapa off-line esse indicador auxilia na seleção do melhor conjunto de hiperparâmetros do modelo. Já na segunda etapa a métrica tem a função de demonstrar como o modelo se comporta com um conjunto de dados não observado em seu treinamento. Desta forma, o erro cometido na última etapa pode ser entendido como o erro esperado quando da aplicação da metodologia no ambiente do COD.

Vale destacar que essa medida apresenta problemas quando há valores reais nulos, ou muito próximos de zero, dado que nesses casos o MAPE tenderá a  $\pm\infty$ , porém não significando necessariamente que o modelo não é adequado.

#### 4.4. Resultados

O método desenvolvido foi aplicado no sistema de 34 barras de distribuição [Soares et al. 2019]. Foram selecionadas as barras 1, 3, 6, 7, 10, 12, 16, 27, 30 e 34 para conter as medidas SCADA. As demais barras possuem medidas AMI, as quais não há acesso em tempo real para execução da estimação de estado. As subseções a seguir apresentam os resultados da seleção da arquitetura dos modelos de AE e a *performance* da reconstrução das medidas.

##### 4.4.1. SELEÇÃO DE ARQUITETURA

Nesta etapa o *framework* apresentado em [Akiba et al. 2019] foi utilizado com o TPE como otimizador de modo a encontrar qual a melhor arquitetura possível para a rede neural. Vale ressaltar que o objetivo foi minimizar o MAPE dos dados de validação previamente separados.

O melhor resultado obtido foi com um modelo com 3 camadas ocultas e com uma taxa de aprendizado de  $\approx 9.06 \cdot 10^{-5}$ . Para cada uma das camadas tem-se a quantidade de neurônios selecionados e a taxa de esquecimento. Vale ressaltar que a primeira e a segunda camada correspondem ao processo de codificação, enquanto a terceira e a última camada a fazem parte da decodificação.

**Tabela 2. Estrutura Selecionada pelo Problema de Otimização**

Camada	Qtd Neurônios	Taxa de Esquecimento
1	224	$\approx 0,025$
2	67	$\approx 0,019$
3	79	$\approx 0,010$



#### 4.4.2. APLICAÇÃO DO MODELO

Após a seleção do modelo foi simulada a sua aplicação de modo a avaliar o desempenho obtido perante o conjunto de dados de teste.

A base de teste possui 20% da quantidade total de dados gerados. Esse conjunto não foi aplicado em nenhuma outra etapa do treinamento. Assim, acredita-se que os resultados obtidos com eles representem o real comportamento do modelo.

As estatísticas obtidas com o modelo são apresentadas nas Tabelas (3) e (4). A Tabela (3) expõe o resultado do primeiro quartil (Q1), mediana, média e terceiro quartil (Q3) do erro percentual absoluto. O valor obtido com a aplicação da Equação (2) corresponde à coluna da média na Tabela (3).

**Tabela 3. Quartis e Média dos Resultados**

Tipo de Medida	Q1	Mediana	Média	Q3
Tensão - AMI	0,0012	0,0031	0,0055	0,0068
Pot. Ativa - AMI	0,9739	2,1317	2,3661	3,4604
Pot. Ativa - SCADA	0,2261	0,5112	0,8151	1,0269
Pot. Reativa - AMI	1,8797	4,6783	458,84	8,4725
Pot. Reativa - SCADA	0,3899	0,9010	223	1,8694

Já a Tabela (4) apresenta as médias de dispersão relacionadas ao erro percentual absoluto, sendo  $\sigma$  o desvio padrão e  $MAD$  desvio em relação à mediana. A coluna  $P90$  representa o percentil 90 em relação aos dados e  $P99$  o percentil 99.

**Tabela 4. Desvio Padrão, Desvio Mediano e Percentis dos Resultados**

Tipo de Medida	P90	P99	$\sigma$	MAD
Tensão - AMI	0,0124	0,0378	0,0078	0,0033
Pot. Ativa - AMI	4,5836	7,2648	1,7718	1,8236
Pot. Ativa - SCADA	1,8539	4,7979	0,9706	0,5087
Pot. Reativa - AMI	12,71	42,39	15261	4,6516
Pot. Reativa - SCADA	3,4597	11,5416	9108	0,9214

Como pode ser observado, as medidas de potência reativa possuem estatísticas de erro muito superiores às demais medidas. Por esse motivo, além das medidas já apresentadas buscou-se verificar algumas situações que poderiam gerar erros elevados.

De modo a identificar situações em que grandes erros na reconstrução das medidas de potência reativa ocorrem, foi investigada a relação entre tais erros e a potência reativa média fornecida pela subestação. A partir disso foi possível definir um valor mínimo de potência reativa média entre às três fases, a ser usado em um filtro para melhorar a confiabilidade do modelo. O valor de 0,3498 foi definido para a construção de tal filtro, em que quando a potência reativa média se encontra acima deste valor os resultados do modelo são melhores. Vale ressaltar que abaixo deste limiar há somente 1,04% dos dados. Assim, por mais que se defina um filtro, não há perda na generalização do modelo. A Tabela (5) apresenta as estatísticas antes da aplicação do filtro, após a sua aplicação e o

fator de redução proporcionado pelo mesmo. Como é possível observar, há uma mudança significativa nas métricas média, desvio padrão e  $P99$ .

**Tabela 5. Mudança nas Estatísticas com a Aplicação do Filtro - Pot. Reativa AMI**

Estatística	Antes	Depois	Fator de Redução
Média	458,84	152,3	67%
$\sigma$	15261,45	4399,04	71%
P99	42,39	32,44	23%

Outro fator relevante é compreender quais medidas AMI reconstruídas possuem maior erro. Tendo em vista que com a reconstrução das medidas AMI há mais medidas que o necessário para garantir a observabilidade da rede pode-se descartar medidas que se sabe que o modelo tem maior dificuldade para reconstruir. Alternativamente, quando usadas pelo estimador de estado tais medidas reconstruídas podem ser usadas como pseudomedidas de baixa qualidade, ou seja, com baixos pesos a ela atribuídos. Para esse indicador foi utilizado o percentil 99 do erro percentual absoluto das pseudomedidas geradas. Nesse caso, a barra 29 destacou-se com os maiores percentis. As demais medidas possuem percentis 99 entre 10% e 20%. Isso significa que os erros mais elevados encontram-se em medidas tomadas na barra 29. O maior erro encontrado em medidas tomadas em outras barras foi em torno de 20%. Haja vista que erros contidos em pseudomedidas colhidas de bases históricas de carga podem variar entre 20% e 50% [Singh et al. 2009], os resultados obtidos indicam que pseudomedidas de muito melhor qualidade podem ser obtidas com o modelo proposto.

Foram também obtidos resultados do modelo considerando o filtro com a exclusão das medidas de potência reativa da barra 29 ainda na fase de treinamento. A Tabela (6) mostra as estatísticas referentes à reconstrução das medidas reativas neste caso (Filtrado) e também para o caso inicial (Bruto), para o qual as estatísticas foram apresentadas nas Tabelas (3) e (4).

**Tabela 6. Comparação da Reconstrução da Potência Reativa: Caso Bruto e Caso Filtrado**

	Q1	Mediana	Média	Q3	P90	P99	$\sigma$	MAD
Bruto	1,8797	4,6783	458,84	8,4725	12,71	42,39	15261	4,6516
Filtrado	1,8571	4,5929	5,6239	8,1618	11,7443	20,47	5,17	4,497
F.R.	1%	2%	99%	4%	8%	52%	100%	3%

Como pode ser observado há uma melhoria significativa nos indicadores quando se considera o caso filtrado.

#### **4.4.3. CONSIDERAÇÕES**

O modelo de AE demonstra uma boa capacidade para a geração de pseudomedidas. Essa boa *performance* pode ser proveniente do fato do modelo tratar todas as medidas e pseudomedidas ao mesmo tempo. Embora a reconstrução da potência reativa continua sendo de qualidade inferior à obtida para as medidas de potência ativa e de tensão, os erros observados apresentam-se na mesma ordem de grandeza (médias inferiores a 5% de MAPE

por exemplo), exceto o percentil 99 (P99) que continua com um valor mais alto, porém aceitável. Os resultados obtidos revelam a importância de se tratar ou filtrar *outliers* que podem comprometer o desempenho do modelo. No estudo realizado neste artigo verificou-se que tais situações estão associadas à reconstrução de medidas de potência reativa que possuem valores muito baixos.

## 5. Conclusão

As redes de distribuição sofrem com a falta de medidas disponíveis. Tendo isso em vista utilizar as infraestruturas avançadas de medição torna-se indispensável de modo a garantir a observabilidade do sistema. Entretanto, a utilização desse tipo de medição é dificultada pela baixa frequência em que estão disponíveis para processamento. Neste artigo é proposta a utilização de pseudomedidas para complementar medidas SCADA nos momentos em que medidas AMI não estão disponíveis. Tais medidas são obtidas em tempo real por uma rede neural do tipo *AutoEncoder*, previamente treinada em um ambiente offline. Testes foram realizados com uma rede de 34 barras e os resultados mostraram que medidas de tensão e potência ativa foram reconstruídas com excelente qualidade. Por outro lado, a reconstrução de potência reativa apresentou erros elevados, os quais foram significativamente reduzidos com o emprego de uma etapa de filtragem que reduz o efeito de *outliers*. O modelo proposto conseguiu gerar pseudomedidas de ótima qualidade para processamento pela função estimação de estado, garantindo também a observabilidade da rede quando apenas medidas SCADA estão disponíveis.

## Referências

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyperparameter optimization. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc.
- Dahale, S., Karimi, H. S., Lai, K., and Natarajan, B. (2020). Sparsity based approaches for distribution grid state estimation—a comparative study. *IEEE Access*, 8:198317–198327.
- Dehghanpour, K., Yuan, Y., Wang, Z., and Bu, F. (2019). A game-theoretic data-driven approach for pseudo-measurement generation in distribution system state estimation. *IEEE Transactions on Smart Grid*, 10(6):5942–5951.
- He, Z., Hong, K., Zhou, J., Liang, D., Wang, Y., and Liu, Q. (2022). Deep frequency-recurrent priors for inverse imaging reconstruction. *Signal Processing*, 190:108320.
- Hinton, G. E. and Zemel, R. (1994). Autoencoders, minimum description length and helmholtz free energy. In Cowan, J., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems*, volume 6. Morgan-Kaufmann.
- Huang, S.-C., Lu, C.-N., and Lo, Y.-L. (2015). Evaluation of ami and scada data synergy for distribution feeder modeling. *IEEE Transactions on Smart Grid*, 6(4):1639–1647.

- Jabbar, A., Li, X., and Omar, B. (2020). A survey on generative adversarial networks: Variants, applications, and training.
- Ko, J. U., Na, K., Oh, J.-S., Kim, J., and Youn, B. D. (2021). A new auto-encoder-based dynamic threshold to reduce false alarm rates for anomaly detection of steam turbines. *Expert Systems with Applications*, page 116094.
- Krstulovic, J., Miranda, V., Costa, A. J. S., and Pereira, J. (2013). Towards an auto-associative topology state estimator. *IEEE transactions on power systems*, 28(3):3311–3318.
- Lefebvre, S., Prévost, J., and Lenoir, L. (2014). Distribution state estimation: A necessary requirement for the smart grid. In *2014 IEEE PES General Meeting — Conference Exposition*, pages 1–5.
- Li, H., Lai, L., and Zhang, W. (2011). Communication requirement for reliable and secure state estimation and control in smart grid. *IEEE Transactions on Smart Grid*, 2(3):476–486.
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26.
- Mestav, K. R., Luengo-Rozas, J., and Tong, L. (2019). Bayesian state estimation for unobservable distribution systems via deep learning. *IEEE Transactions on Power Systems*, 34(6):4910–4920.
- Miranda, V., Krstulovic, J., Keko, H., Moreira, C., and Pereira, J. (2011). Reconstructing missing data in state estimation with autoencoders. *IEEE Transactions on power systems*, 27(2):604–611.
- Singh, R., Pal, B., and Jabr, R. (2009). Choice of estimator for distribution system state estimation. *IET generation, transmission & distribution*, 3(7):666–678.
- Soares, W., de Souza, J. C. S., Do Coutto Filho, M. B., and Augusto, A. A. (2019). Distribution system state estimation with real-time pseudo-measurements. In *2019 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*, pages 1–5. IEEE.
- Vijayalakshmi, G. and Shanthakumar, M. (2019). Image restoration on fusion of mammographs and mri breast images using dual tree complex wavelet transform. *International Journal of Advanced Science and Technology*, 28(17):842–857. cited By 0.
- Xu, W., Zhang, J., Li, X., Yuan, S., Ma, G., Xue, Z., Jing, X., and Cao, J. (2022). Intelligent denoise laser ultrasonic imaging for inspection of selective laser melting components with rough surface. *NDT & E International*, 125:102548.
- Yilmaz, C. S., Yilmaz, V., and Gungor, O. (2022). A theoretical and practical survey of image fusion methods for multispectral pansharpening. *Information Fusion*, 79:1–43.
- Zhang, X., Mu, J., Zhang, X., Liu, H., Zong, L., and Li, Y. (2022). Deep anomaly detection with self-supervised learning and adversarial training. *Pattern Recognition*, 121:108234.
- Zhao, J., Huang, C., Mili, L., Zhang, Y., and Min, L. (2020). Robust medium-voltage distribution system state estimation using multi-source data. In *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5.