

Validação de dados de saída de um modelo de Simulação a Eventos Discretos utilizando Redes Neurais Adversárias



Gustavo Teodoro Gabriel (Universidade Federal de Itajubá)
gustavo.teodoro.gabriel@gmail.com

Afonso Teberga Campos (Universidade Federal de Itajubá)
teberga@unifei.edu.br

José Arnaldo Barra Montevechi (Universidade Federal de Itajubá)
montevechi@unifei.edu.br

Fabiano Leal (Universidade Federal de Itajubá)
fleal@unifei.edu.br

A validação de modelos computacionais de Simulação a Eventos Discretos (SED) é primordial para o sucesso do projeto, uma vez que é a partir dessa etapa que se garante que o modelo simulado corresponde ao sistema real. Como não é possível garantir que o modelo represente 100% do sistema real, é necessário desenvolver técnicas para chegar ao mais próximo possível. Apesar da literatura apresentar uma gama de técnicas, sugere-se que use uma abordagem estatística na validação. Apesar disso, modelos em que há alta variabilidade dos processos podem não se enquadrar nos pressupostos necessários. Assim, o objetivo do trabalho é verificar a aplicabilidade das Redes Neurais Adversárias, ou seja, as Generative Adversarial Networks (GANs) para validar modelos de simulação. Para isso, foi desenvolvido um passo a passo para a validação de modelos computacionais utilizando as GANs. Para o teste, foi desenvolvido e programado o treinamento das GANs no Python para a distribuição normal e triangular. Após o treinamento e as GANs cumprirem os requisitos necessários, as distribuições consideradas reais foram comparadas a outras distribuições, variando os parâmetros de cada uma delas. Com o teste de equivalência foi possível observar a eficiência das GANs ao julgar dados.

Palavras-chave: Simulação a eventos discretos, GANs, validação computacional, Teste de Equivalência.

1. Introdução

Modelos de Simulação a Eventos Discretos (SED) surgiram na área militar e desde então são utilizados para a implementação de novos sistemas e melhorias de processo, análise de capacidade, alocação de recursos, auxílio na tomada de decisão, entre outros. Scheidegger et al. (2018) identificaram mais de 20 áreas nas quais a SED foi utilizada, como manufatura, transporte, serviços e saúde. Além disso, com o avanço da indústria 4.0, os modelos vêm se tornando cada vez mais sofisticados e sendo utilizados em tempo real ou quase-real (RODIČ, 2017).

Apesar desse crescente avanço, os modelos só trazem resultados efetivos se correspondem de fato ao sistema simulado. Para isso, a etapa de validação computacional, presente em diversos métodos de simulação, é essencial para garantir que o modelo simulado represente o sistema real, uma vez que o processo de simulação é propenso a erros (HARREL et al., 2012). Além disso, se o modelo construído não refletir a realidade, os experimentos realizados no modelo de simulação não darão uma resposta efetiva para o sistema real (BANKS e CHWIF, 2011).

Dessa forma, Zeigler e Nutaro (2016) destacam que a validade é o grau em que um modelo representa fielmente sua contraparte do sistema. Os processos de validação têm como objetivo a quantificação da precisão do modelo, comparando os resultados da simulação com resultados experimentais ou operacionais no mundo real (THACKER et al., 2004), determinando se o modelo representa adequadamente o sistema para o propósito desejado (LAW, 2015; SARGENT e BALCI, 2017). Com isso, mesmo que não se possa garantir que um modelo seja 100% válido, a validação é parte crucial para que os resultados obtidos do modelo de simulação sejam confiáveis (CHWIF, MUNIZ e SHIMADA, 2008; OLSEN e RAUNAK, 2015).

Todas as métricas de saída relacionadas aos objetivos do problema devem ser validadas. A literatura apresenta diversas técnicas de validação, apesar disso, Sargent (2013) afirma que comparações estatísticas entre o modelo simulado e o sistema real são preferíveis por apresentar uma abordagem objetiva e evidências matemáticas. No entanto, em se tratando de modelos com alta variabilidade, os dados podem ser correlacionados automaticamente ou premissas de normalidade e tamanho de amostra podem ser violadas, impedindo o uso de técnicas estatísticas clássicas (COMAS et al., 2008). Para resolver o problema pode-se utilizar as *Generative Adversarial Networks* (GANs), ou seja, Redes Adversárias Generativas. Elas foram desenvolvidas por Goodfellow et al. (2014) e são caracterizadas como modelos de

aprendizagem profunda. Nesse sentido, o artigo tem como objetivo propor um método de validação para modelos de SED e mostrar o funcionamento das GANs na validação de dados. O artigo está estruturado da seguinte maneira: a seção 1 apresentou a introdução e a contextualização do problema; a seção 2 apresenta a revisão bibliográfica; logo em seguida, na seção 3, é apresentado o método proposto; os resultados e discussões estão na seção 4 e por fim, na seção 5 estão as conclusões.

2. Referencial Teórico

2.1. Validação de modelos computacionais

Validade é considerada como o grau em que um modelo representa fielmente sua contraparte do sistema (ZEIGLER e NUTARO, 2016). Sua avaliação é de suma importância para projetos de simulação, visto que, por natureza, o processo de construção de um modelo é propenso a erros, gerados no processo de tradução do sistema real em um modelo conceitual que será, mais tarde, convertido em um modelo computacional (HARREL et al., 2012).

Nesse sentido, processos de verificação e validação (V&V) foram desenvolvidos para minimizar esses erros e fazer modelos úteis para atender às finalidades para as quais foram criados (LEAL et al., 2011). Enquanto a verificação do modelo visa garantir que as lógicas do modelo computacional e sua implementação estejam corretas, a validação do modelo tem como objetivo a comprovação de que um modelo computacional possui uma faixa satisfatória de precisão consistente com o propósito pretendido sobre o domínio da aplicabilidade planejada (SARGENT, GOLDSMAN e YAACUB, 2016).

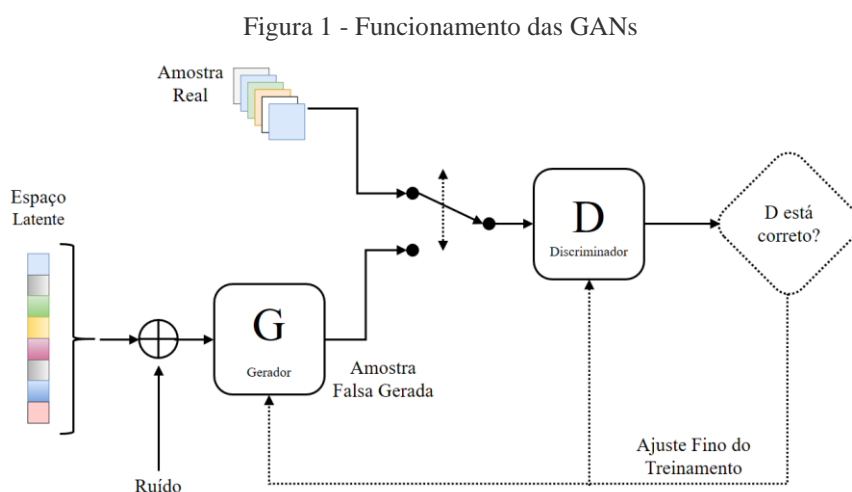
Ou seja, os processos de validação têm como objetivo a quantificação da precisão do modelo, comparando os resultados da simulação com resultados experimentais ou operacionais no mundo real (THACKER et al., 2004), determinando se o modelo representa adequadamente o sistema para o propósito desejado (LAW, 2015; SARGENT e BALCI, 2017).

2.2. Redes Neurais Adversárias

As Redes Adversárias Generativas, que vem do inglês Generative Adversarial Networks (GANs) foram propostas por Goodfellow et al. (2014) e são caracterizadas como modelos de aprendizagem profunda. Elas são estudadas e utilizadas para gerar imagens falsas, ou também chamadas de sintéticas, mas, significativamente realistas que tem impactado positivamente a área de visão computacional. Atualmente, as GANs recebem atenção por serem empregadas para a geração de manipulação de imagens, vídeos e sons, com o objetivo de gerar mídias falsas

(SORIN et al., 2020). Karras et al. (2019) mostram em seu estudo que é possível encontrar páginas onde se criam fotos aleatórias de pessoas que na verdade não existem, porém são geradas a partir de GANs.

Segundo Pan et al. (2019), as GANs foram criadas a partir da teoria dos jogos, onde duas redes neurais artificiais competem entre si, sendo uma gerando dados, enquanto a outra é considerada discriminadora. A primeira, ou seja, a geradora, tem como objetivo a geração de dados e amostras falsas para enganar a rede discriminadora. A segunda, por sua vez, tem o papel de diferenciar se os dados gerados são falsos ou verdadeiros. Com o passar do tempo, o aprendizado e o treinamento das GANs, as redes geradoras começam a produzir amostras que são similares às amostras reais. Portanto, essa estrutura permite gerar novas amostras sintéticas, mas que se enquadram na população verdadeira da qual foi gerada, caracterizando um método Aprendizagem não Supervisionada (BROWNLEE, 2020). A Figura 1 mostra como é o funcionamento das GANs.



Fonte: Pan et al. (2019)

3. Método Proposto

O método proposto para a validação dos dados se inicia com a definição dos parâmetros de entrada para o treinamento das GANs. Após análises da literatura e testes de treinamento, conclui-se que é necessário um número de amostras grande para que o algoritmo entenda a classificação dos dados. Além disso, o número de camadas e neurônios devem ser maiores para o discriminador do que para o gerador. O mesmo é válido para a taxa de aprendizado, pois o discriminador deve aprender a discriminar os dados com mais facilidade que o gerador. Essa taxa é compensada ao escolher um decaimento do aprendizado com valor entre 0,8 e 0,9.

Logo em seguida, os dados de referência, ou seja, aqueles que representam o sistema real são inseridos na programação e servem de base para o treinamento das GANs. Todos os dados são escalonados para um intervalo de $[-1,1]$ pois facilita a geração e desenvolvimento de todo o algoritmo, apresentando resultados melhores em menos tempo e iterações.

A cada época, o discriminador e o gerador são treinados. O gerador gera amostras aleatórias com ruídos e os dados tanto reais quanto sintéticos são divididos em dois lotes e enviados para o discriminador. O discriminador, a cada iteração, mede a acurácia, classificando os dados como verdadeiros (1) ou falsos (0). A acurácia, ou seja, a probabilidade de os dados verdadeiros serem verdadeiros (p_{vv}) é medida como a porcentagem de classificação, conforme mostra a Equação 1. Também é computado a função perda de cada um.

$$p_{vv} = \frac{\sum_{i=1}^{n_r} p_i}{n_r} \quad (1)$$

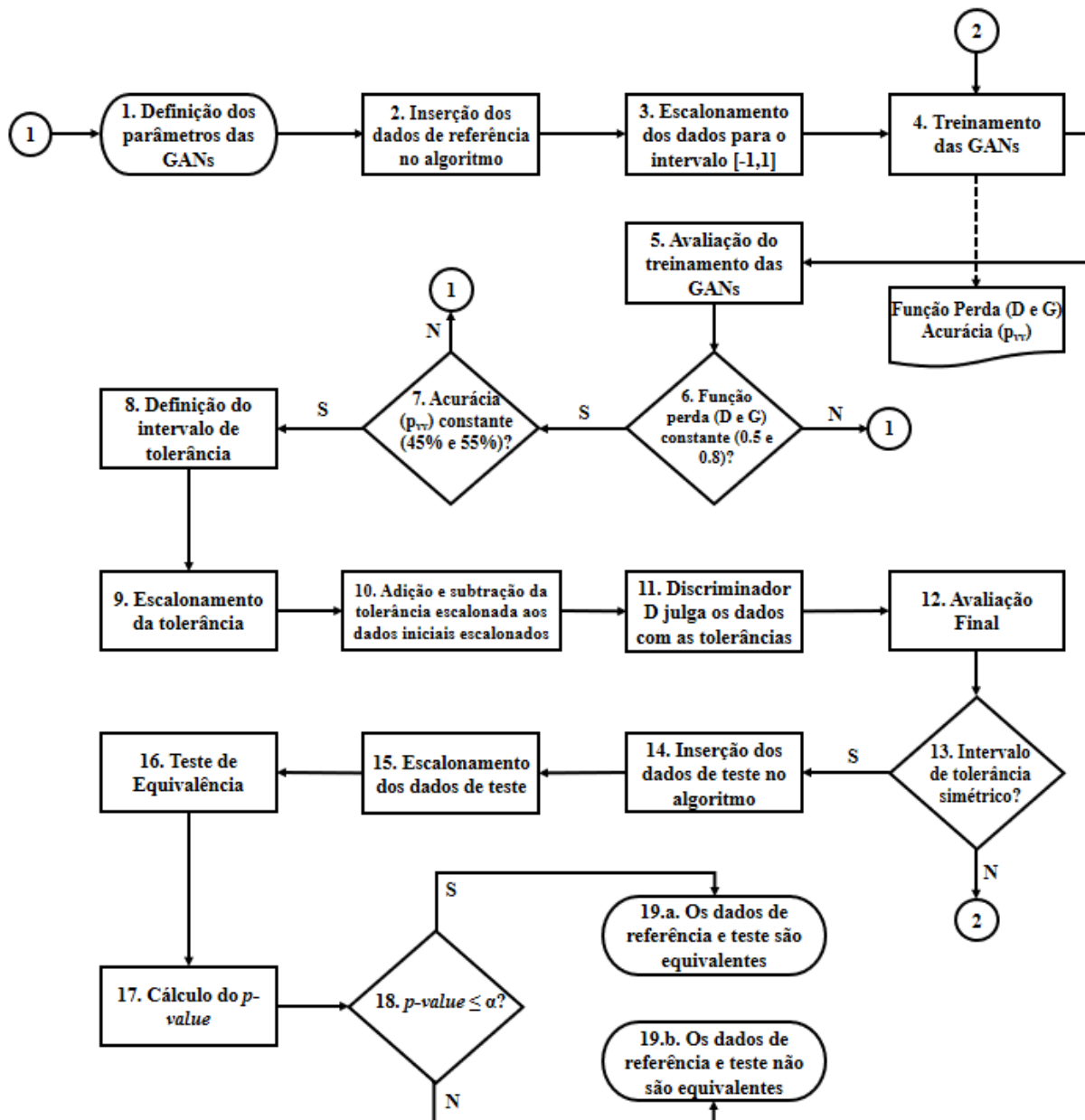
As GANs apresentam bons resultados se o discriminador avalia os dados entre 45,0% e 55,0% e se mantém constante ao longo das últimas iterações. Ainda, as funções de perda devem estar estabilizadas entre os valores 0,5 e 0,8. Caso esses critérios não sejam atendidos, é necessário reavaliar os parâmetros iniciais.

Após o treinamento, é necessário definir o intervalo no qual os dados de teste são considerados válidos. Dessa maneira, a tolerância definida no início do processo é também escalonada para o intervalo de $[-1,1]$. A tolerância escalonada é somada aos dados de referência escalonados que foram utilizados para o treinamento das GANs. Em seguida, esses dados somados à tolerância escalonada são submetidos a avaliação do discriminador que diz qual a probabilidade de os dados verdadeiros serem verdadeiros. O processo é repetido, porém subtraindo a tolerância escalonada dos dados de referência.

Nesse ponto há a última avaliação da eficiência das GANs. Ao avaliar os dados de referência com a soma e a subtração da tolerância, o discriminador deve apresentar um intervalo simétrico (se a tolerância definida for simétrica). Ou seja, ao julgar os dados com as tolerâncias, ele deve apresentar um intervalo em que os dados reais se encontram no centro. O limite inferior do intervalo é dado por $\overline{p_{r,tl}}$, enquanto o limite superior é dado por $\overline{p_{r,ts}}$. Se o limite não for simétrico significa que as GANs não são eficientes para a discriminação dos dados, mas podem ser eficientes para a geração. Caso esse critério de avaliação não seja atendido, é necessário refazer o treinamento, mas não é necessário redefinir os parâmetros.

Após essa avaliação, pode-se medir os dados de teste. Eles são inseridos e também convertidos na mesma escala dos dados de referência. Aqui, os dados são convertidos em uma escala levando em consideração o primeiro escalonamento. Em seguida, é possível realizar o Teste de Equivalência (SCHUIRMANN, 1981) para julgar se a diferença entre os dados está no intervalo determinado. A Figura 2 mostra o fluxograma do método proposto.

Figura 2 - Passo a passo para a validação dos dados através das GANs e Teste de Equivalência



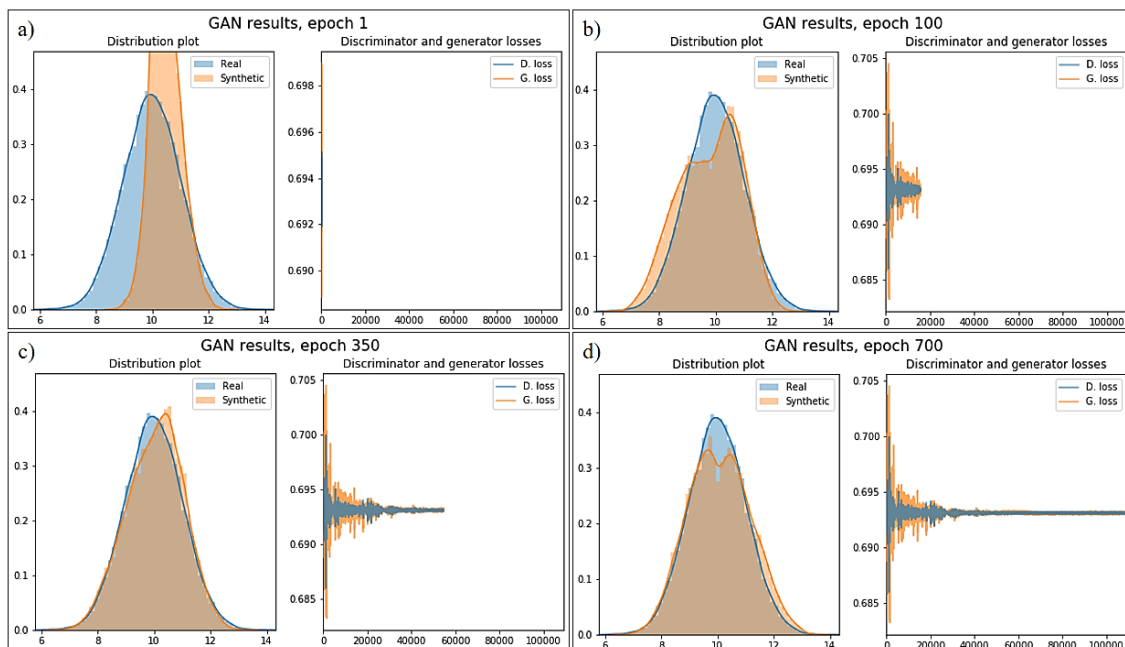
4. Análise e Resultados

Os testes foram realizados para distribuições estatísticas vastamente exploradas na literatura, as quais pode-se ter o controle dos dados. Foram escolhidas as distribuições Normal e Triangular para mostrar o procedimento descrito na seção 3.

4.1. Distribuição normal

Foi gerada uma distribuição $N(10; 1)$ com 20000 dados, considerados como os dados de referência e foram utilizados para o treinamento das GANs. Definiu-se um limite de tolerância de 5% para mais e para menos na média dos dados para considerar valores aceitáveis ao fazer o teste. Dessa maneira, uma distribuição normal entre $N(9,5; 1)$ e $N(10,5; 5)$ deveria ser considerada válida ao realizar o Teste de Equivalência. Utilizou-se 700 épocas pois observou-se que a acurácia se estabiliza a partir da época 500. A Figura 3 mostra a evolução da função perda do discriminador e do gerador ao longo das 700 iterações e o comportamento dos dados gerados pelo gerador em 4 épocas diferentes.

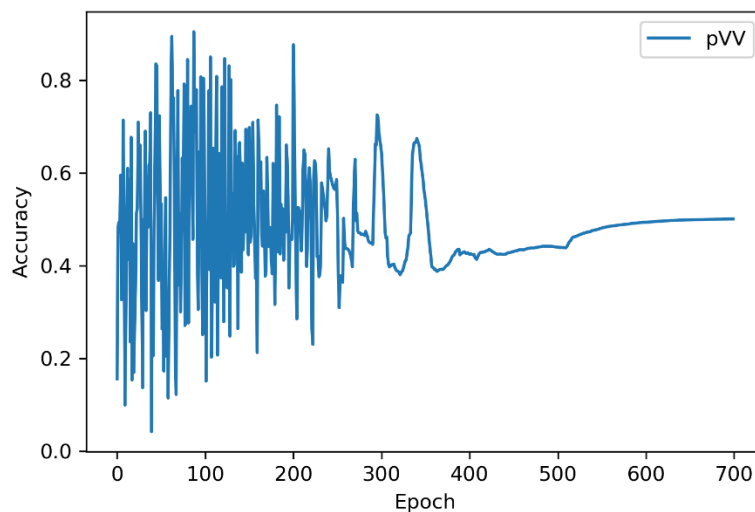
Figura 3 - Evolução do treinamento das GANs – Normal



É possível observar que a função perda do discriminador e do gerador oscila nas primeiras iterações e, a partir da época 350 começa a se tronar constante. O valor da função perda de ambos gira em torno de 0,695, o que está de acordo com a literatura e indica que não houve colapso no treinamento. Apesar dos dados gerados na última época se encontrarem um pouco

abaixo dos dados reais, a acurácia (p_{VV}) dos dados terminou constante no valor de 50,09% ($\overline{p_r}$), conforme a Figura 4. Ou seja, os dados sintéticos, ao serem julgados pelo discriminador estão ao acaso. Isso significa que o gerador conseguiu enganar o discriminador e este não sabe distinguir dados reais de dados sintéticos. Assim, o treinamento passou pelos dois primeiros pontos de avaliação (atividades 6 e 7) do fluxograma apresentado na Figura 2.

Figura 4 - Evolução da acurácia do discriminador ao longo das épocas – Normal



Após a validação dos dois primeiros pontos, os dados de referência somados a tolerância foram submetidos à avaliação do discriminador. O mesmo foi feito subtraindo a tolerância dos dados de referência. Ao fazer essa avaliação, obteve-se o valor de 39,13% para o limite superior da tolerância e 60,245% para a avaliação do limite inferior. Dessa maneira, tem-se que $\overline{p_{r,ti}} = 0,3913$ e $\overline{p_{r,ts}} = 0,60245$. Conclui-se assim, que os limites de equivalência são os valores de $\overline{p_{r,ti}}$ e $\overline{p_{r,ts}}$. Portanto, ao se avaliar os dados de teste, o discriminador deve julgá-los dentro desse limite. Se o discriminador julgar fora, significa que os dados não podem ser considerados iguais aos dados de referência, ou seja, iguais aos dados reais.

Como o ponto de avaliação 13 da Figura 2 foi atendido, foi possível passar para a fase de experimentação para provar que as GANs podem julgar os dados e ser útil na validação. Assim, foram realizados 144 testes, variando a média da distribuição, o desvio padrão e o número de amostra em 9, 3 e 4 níveis respectivamente. Os níveis testados estão apresentados na Tabela 1. Como os valores são conhecidos, espera-se que as distribuições com médias inferiores à 9,5 e superiores à 10,5 não estejam dentro do intervalo.

Tabela 1 - Níveis das variáveis de teste - Normal

Fator	Níveis									
Média	9.00	9.25	9.50	9.75	10.00	10.25	10.50	10.75	11.00	
Desvio Padrão	0.5	1.0	1.5							
Número de amostras	10	100	1000	10000						

Assim, o Teste de Equivalência é dado por:

$$H01: \bar{p}_r - 0.5009 < 0.3913 \text{ ou } H02: \bar{p}_r - 0.5009 > 0.60245$$

$$H1: 0.3913 \leq \bar{p}_r - 0.5009 \leq 0.60245$$

Os resultados encontrados para os testes estão resumidos na Tabela 2 (*p-value*) e 3 (Poder do teste).

Tabela 2 - Resultado dos Testes de Equivalência (*p-value*) - Normal

<i>p-value</i> _(máx)		Média								
S _t	n _t	9.00	9.25	9.50	9.75	10.00	10.25	10.50	10.75	11.00
0.5	10	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000
	10000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000
1.0	10	1.000	0.989	0.993	0.000	0.000	0.000	0.995	0.665	0.986
	100	1.000	1.000	1.000	0.000	0.000	0.000	1.000	1.000	1.000
	1000	1.000	1.000	1.000	0.000	0.000	0.000	1.000	1.000	1.000
	10000	1.000	1.000	1.000	0.000	0.000	0.000	1.000	1.000	1.000
1.5	10	1.000	0.628	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	100	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	10000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

p-value $\geq \alpha$, ou seja, não há equivalência entre os dados de teste e referência

Como o julgamento do discriminador leva em consideração o comportamento dos dados em toda a distribuição, pode-se concluir que, ao analisar as distribuições com desvio padrão menor (0,5) que a distribuição original, todos os testes de equivalência foram rejeitados, com exceção da distribuição com média 10. Por outro lado, o discriminador julgou os dados iguais ao discriminar as distribuições que apresentam desvio padrão maior (1,5) que a distribuição original. Isso se deve ao fato de que os dados se apresentam mais espaçados e as GANs levam esse espaçamento em consideração.

Tabela 3 - Resultado dos Testes de Equivalência (Poder do Teste) – Normal

Poder		Média								
S_t	n_t	9.00	9.25	9.50	9.75	10.00	10.25	10.50	10.75	11.00
0.5	10	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	100	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	1000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	10000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
1	10	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.769	0.000
	100	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
	1000	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
	10000	0.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
1.5	10	0.000	0.830	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Quando se trata das distribuições com desvio padrão 1, observou-se que o discriminador conseguiu julgar corretamente, ou seja, como equivalentes as distribuições com médias superiores a 9,5 e inferiores a 10,5. Apesar disso, os limites das tolerâncias se apresentaram mais sensíveis as análises. Para o limite superior, ao testar uma distribuição N (10,50; 1) os dados não foram considerados válidos ($p\text{-value} = 1,000$), porém para uma distribuição N (10,49; 1) os dados podem ser considerados equivalentes ($p\text{-value} = 0,000$). O mesmo é válido para o limite inferior, onde a distribuição N (9,50; 1) não é considerada equivalente aos dados de referência ($p\text{-value} = 1,000$), mas a distribuição N (9,53; 1) é ($p\text{-value} = 0,0000$).

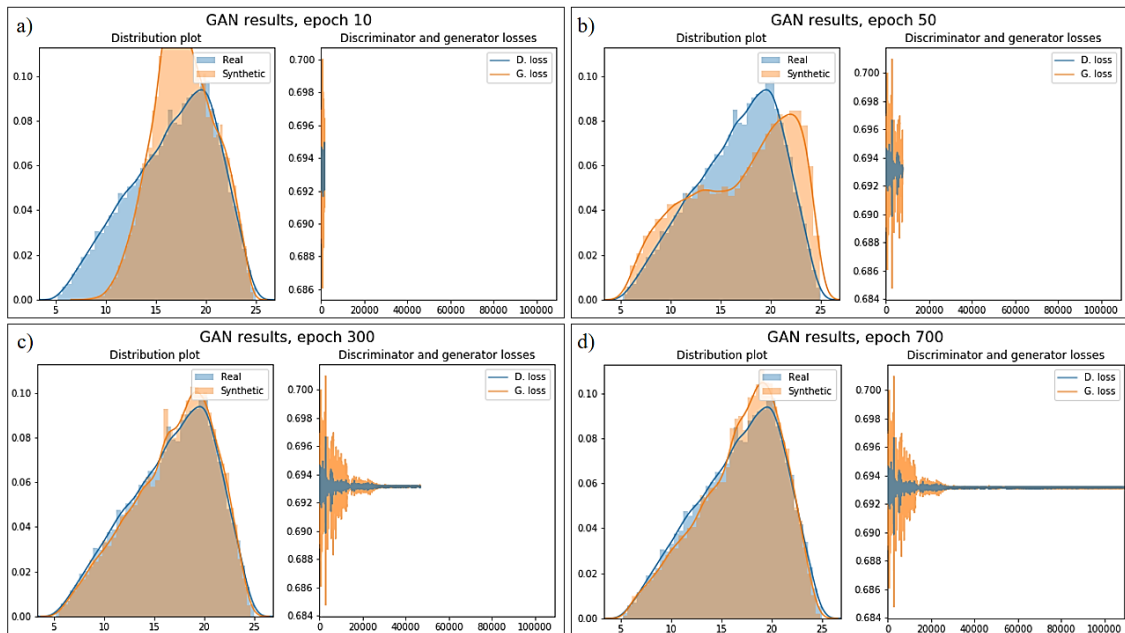
Em relação ao número de amostras testadas, é possível perceber que o discriminador consegue julgar os dados, mesmo apresentando uma amostra pequena de teste. O número de amostra influencia no poder do teste, conforme explicado por Walker e Nowacki (2011) e Meyners (2012). Quanto maior o número de dados na amostra, maior o poder do teste. Além disso, quando a diferença entre as duas médias se aproxima de zero, o poder é maior.

4.2. Distribuição triangular

A segunda distribuição foi uma Triangular. Utilizou-se como dados reais e para o treinamento das GANs uma distribuição T(5, 20, 25) com 20000 dados. Ou seja, a distribuição apresenta como menor valor 5, como valor mais provável 20 e valor máximo de 25. Da mesma forma que a distribuição normal, definiu-se uma tolerância de 5% em relação a moda da distribuição. Portanto, os valores de teste deveriam estar entre as distribuições T(4, 19, 24) e T(6, 21, 26).

Também foi utilizado 700 épocas para o treinamento. A Figura 5 mostra a evolução da função perda do discriminador e do gerador e os dados sintéticos gerados nas épocas 10, 50, 300 e 700.

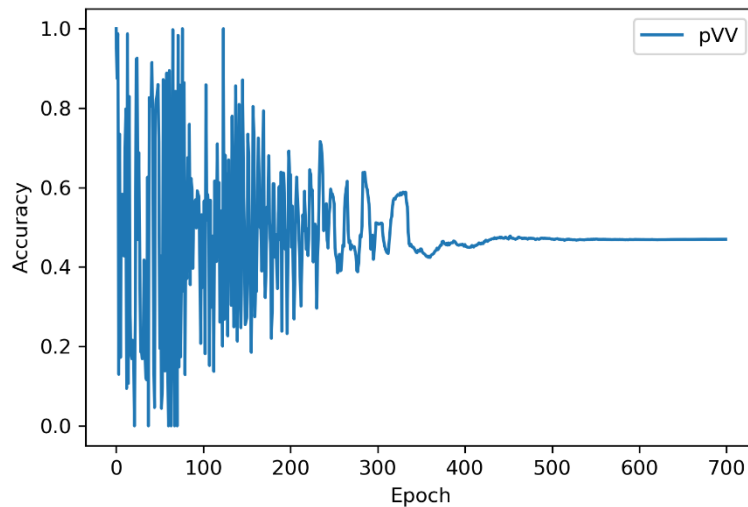
Figura 5 - Evolução do treinamento das GANs - Triangular



A função perda do discriminador e do gerador oscila nas primeiras iterações, sendo que há maior oscilação do gerador que começa se estabilizar e tornar constante a partir da época 200. O valor da função perda, tanto do gerador quanto do discriminador, ao ficar constante está próximo de 0,693. Assim como no treinamento para a distribuição normal, indica que não houve colapso no treinamento. A acurácia (p_{vv}) dos dados gerados pelo gerador terminou constante no valor de 46,985% ($\overline{p_r}$), conforme a Figura 6. De acordo com o método proposto, o valor final de ($\overline{p_r}$) devem estar entre 45% e 55%. Assim, os dados sintéticos, ao serem julgados pelo discriminador estão ao acaso, com uma diferença de 3,105 em relação ao valor ideal. Apesar dessa pequena diferença, o treinamento pode ser considerado válido e indica que passou pelos dois primeiros pontos de avaliação (atividades 6 e 7).

Seguindo o passo a passo da Figura 2, a tolerância em relação a moda foi aplicada nos dados reais e julgada pelo discriminador. Ao subtrair a tolerância dos dados reais, o discriminador atribuiu o valor de 45.07%. Ao somar a tolerância, obteve-se o valor de 47.995%. Dessa maneira, o ponto de avaliação do intervalo de julgamento foi atendido, pois existe um intervalo simétrico.

Figura 6 - Evolução da acurácia do discriminador ao longo das épocas – Triangular



Após essa avaliação, foi criado um teste variando os três parâmetros da distribuição para fazer o julgamento pelo discriminador. Nessa etapa não houve a alteração do número de amostras, uma vez que os testes realizados para a distribuição normal mostraram que as GANs são capazes de fazer o julgamento independentemente dessa quantidade. Foram realizados 125 testes e os níveis testados de cada parâmetro estão apresentados na Tabela 4.

Tabela 4 - Níveis das variáveis de teste - Triangular

Fator	Níveis				
Mínimo	3	4	5	6	7
Moda	18	19	20	21	22
Máximo	23	24	25	26	27

Assim, o Teste de Equivalência é dado por:

$$H01: \bar{p}_r - 0.46985 < 0.4507 \text{ ou } H02: \bar{p}_r - 0.46985 > 0.47995$$

$$H1: 0.4507 \leq \bar{p}_r - 0.46985 \leq 0.47995$$

Os testes para os 125 experimentos estão apresentados na Tabela 5, mostrando o *p-value* máximo de cada rodada.

Tabela 5 - Resultado dos Testes de Equivalência (p-value) - Triangular

<i>p-value</i> (máx)		Moda				
Mínimo	Máximo	18	19	20	21	22
3	23	1.000	0.000	0.000	1.000	0.000
	24	1.000	1.000	0.000	0.000	1.000
	25	1.000	1.000	0.000	0.000	1.000
	26	1.000	1.000	0.000	0.000	1.000
	27	1.000	0.000	0.000	0.000	0.000
4	23	1.000	0.000	0.000	1.000	0.000
	24	1.000	1.000	0.000	0.000	1.000
	25	1.000	1.000	0.000	0.000	1.000
	26	1.000	0.000	0.000	0.000	0.000
	27	1.000	0.000	0.000	0.000	0.000
5	23	1.000	0.000	1.000	1.000	0.000
	24	1.000	0.000	0.000	0.000	1.000
	25	1.000	0.000	0.000	0.000	1.000
	26	1.000	0.000	0.000	0.000	0.000
	27	0.000	0.000	0.000	1.000	0.000
6	23	0.000	0.000	1.000	1.000	1.000
	24	1.000	0.000	1.000	1.000	0.000
	25	0.000	0.000	1.000	0.000	0.000
	26	0.000	0.000	1.000	1.000	0.000
	27	0.000	1.000	1.000	1.000	1.000
7	23	0.000	1.000	1.000	1.000	1.000
	24	0.000	1.000	1.000	1.000	0.000
	25	0.000	1.000	1.000	1.000	0.000
	26	0.000	1.000	1.000	1.000	1.000
	27	1.000	1.000	1.000	1.000	1.000

De acordo com a Tabela 5 quando a moda se afasta do valor da distribuição original e vai se aproximando da tolerância, a distribuição de teste tende a ser rejeitada. O mesmo vale para os valores de mínimo e máximo. Em compensação, se os valores se encontram entre a tolerância estimada, o discriminador tende a aceitar a distribuição dos dados, o que de fato deveria ocorrer. Vale ressaltar que assim como ocorrido na distribuição normal, quando há a aproximação dos limites de tolerância, a avaliação do discriminador começa a se tornar sensível, como é observado para a rejeição das distribuições de moda 20 e o valor mínimo de 6. Esperava-se que os valores máximos entre 24 e 26, o discriminador julgasse como valores iguais aos da distribuição de referência.

Para encontrar esse limite de tolerância aceitável pela distribuição, testou-se algumas combinações. Como resultado final, observou-se que, ao discriminar a distribuição T (3.8, 20, 25) o discriminador a avalia como valores equivalentes. O mesmo vale para a distribuição T (5.8, 20, 25.6). Valores acima de 25.6, o discriminador julga a distribuição de teste diferente da distribuição de referência.

Um último teste foi realizado, no qual testou-se as distribuições T (0, 20, 30) e T (10, 20, 20). O *p-value* de ambas as distribuições foram de 0.000, confirmando que não há equivalência entre as distribuições de teste e de referência. Isso mostra que as GANs julgam o comportamento dos dados ao longo da distribuição.

5. Conclusões

Modelos de SED que apresentam alta variabilidade dos processos são mais difíceis para serem validados. A literatura sugere, sempre que possível, o uso de testes estatísticos. Apesar disso, muitas vezes os dados de saída podem não se enquadrar aos pressupostos necessários para realizar os testes. Uma maneira de sanar esse problema é o uso das GANs. Elas se adaptam a qualquer tipo de dados através de duas redes que competem entre si: o gerador que gera dados e devem ser julgados pelo discriminador.

Através de um método proposto, o teste foi realizado em distribuições teóricas. Duas distribuições, Normal e Triangular, foram treinadas por 700 épocas. Ambos os treinamentos atingiram os requisitos necessários, sendo que a função perda do gerador e discriminador ficou em torno de 0,5 e 0,8, a acurácia do discriminador em torno de 50% e o intervalo de julgamento simétrico.

Foi possível observar que as GANs foram eficientes ao julgar essas distribuições, inclusive com uma tolerância. Isso pode ser visto através do Teste de Equivalência que compara se as distribuições de referência e teste estão dentro de um intervalo.

Como limitação, observou-se a sensibilidade no julgamento do discriminador ao se aproximar do limite de tolerância. Quanto mais próximo, mais sensível se torna e, às vezes, o discriminador pode avaliar um conjunto de dados fora do limite de tolerância sendo que ele deve estar próximo. Além disso, o método é mais vantajoso que os métodos tradicionais uma vez que se adapta a qualquer tipo de dado.

6. Agradecimentos

Os autores agradecem à FAPEMIG, à CAPES, ao CNPq pelo apoio e suporte ao longo da pesquisa.

REFERÊNCIAS

- BANKS, J.; CHWIF, L. Warnings about simulation. **Journal of Simulation**, v.5, n.4, p.279-291, 2011.
- BROWNLEE, J. **Generative Adversarial Networks with Python: Deep Learning Generative Models for Image Synthesis and Image**. 1 ed. Machine Learning Mastery, 2020.
- CHWIF, L.; MUNIZ, P. S.; SHIMADA, L. M. A prescriptive technique for V&V of simulation models when no real-life data are available: results from a real-life project. **Journal of Simulation**, v.2, n.2, p.81-89, 2008.
- COMAS, M.; CASTELLS, X.; HOFFMEISTER, L.; ROMÁN, R.; COTS, F.; MAR, J.; GUTIÉRREZ-MORENO, S.; ESPALLARGUES, M. Discrete-event simulation applied to analysis of waiting lists. Evaluation of a prioritization system for cataract surgery. **Value Health**, v.11, n.7, p.1203-1213, 2008.
- GOODFELLOW, I. J.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. Generative Adversarial Nets. **Advances in Neural Information Processing Systems (NIPS)**, p.1-9, 2014.
- HARRELL, C.; GHOSH, B. K.; BOWDEN, R. **Simulation using ProModel**. 3. ed. New York: McGraw-Hill Education, 2012.
- KARRAS, T.; LAINE, S.; AILA, T. A style-based generator architecture for generative adversarial networks. **In: IEEE Conference on Computer Vision and Pattern Recognition, Proceedings...** Long Beach, CA, USA, p.4396-4405, 2019.
- LAW, A. **Simulation modeling and analysis**. Boston: McGraw-Hill Education, 2015.
- LEAL, F.; COSTA, R. F. S.; MONTEVECHI, J. A. B.; DE ALMEIDA, D. A.; MARINS, F. A. S. A practical guide for operational validation of discrete simulation models. **Pesquisa Operacional**, v.31, n.1, p.57-77, 2011.
- OLSEN, M.; RAUNAK, M. A method for quantified confidence of DEVS validation. **In: Symposium on Theory of Modeling & Simulation: DEVS Integrative M&S Symposium, Proceedings...** Alexandria, VA, USA, p.135-142, 2015.

PAN, Z.; YU, W.; YI, X.; KHAN, A.; YUAN, F.; ZHENG, Y. Recent Progress on Generative Adversarial Networks (GANs): A Survey. **IEEE Access**, v.7, p.36322-36333, 2019.

RODIČ, B. Industry 4.0 and the New Simulation Modelling Paradigm. **Organizacija**, v.50, p.193–207, 2017.

SARGENT, R. G. Verification and validation of simulation models. **Journal of Simulation**, v.7, n.1, p.12-24, 2013.

SARGENT, R. G.; BALCI, O. History of verification and validation of simulation models. **In: Winter Simulation Conference, Proceedings...** Las Vegas, NV, USA, p.292-307, 2017.

SARGENT, R. G.; GOLDSMAN, D. M.; YAACOUB, T. A tutorial on the operational validation of simulation models. **In: Winter Simulation Conference, Proceedings...** Arlington, VA, USA, p.163-177, 2016.

SCHEIDEGGER, A. P. G.; PEREIRA, T. F.; OLIVEIRA, M. L. M.; BANERJEE, A.; MONTEVECHI, J. A. B. An introductory guide for hybrid simulation modelers on the primary simulation methods in industrial engineering identified through a systematic review of the literature. **Computers & Industrial Engineering**, v.124, p.474-492, 2018.

SCHUIRMANN, D. J. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. **Journal of Pharmacokinetics and Biopharmaceutics**, v.15, n.6, p.657-680, 1987.

SORIN, V.; BARASH, Y.; KONEN, E.; KLANG, E. Creating artificial images for radiology applications using Generative Adversarial Networks (GANs) – A Systematic Review. **Academic Radiology**, v.27, n.8, p.1175-1185, 2020.

THACKER, B. H.; DOEBLING, S. W.; HEMEZ, F. M.; ANDERSON, M. C.; PEPIN, J. E.; RODRIGUEZ, E. A. **Concepts of Model Verification and Validation**. Disponível em:
https://inis.iaea.org/collection/NCLCollectionStore/_Public/36/030/36030870.pdf?r=1. 2004

ZEIGLER, B. P.; NUTARO, J. J. Towards a framework for more robust validation and verification of simulation models for systems of systems. **Journal of Defense Modeling and Simulation**, v.13, n.1, p.3-16, 2016.