



DESENVOLVIMENTO DE PROTÓTIPO VIRTUAL DE CÂMARA DE VÁCUO COM *MACHINE LEARNING*

Rodrigo de Souza^{1*} , Francisco Tadeu Degasperis^{1,2} , Marcelo Duduchi Feitosa¹

1. Centro Paula Souza – Coordenadoria Geral de Pós-Graduação, Extensão e Pesquisa – São Paulo (SP), Brasil.

2. Faculdade de Tecnologia de São Paulo – Unidade de Pós-Graduação, Extensão e Pesquisa – São Paulo (SP), Brasil.

Autor correspondente: rodrigosoza.etec@gmail.com

Editora de seção: Mariana Fraga

Recebido: 10 nov. 2025 **Aprovado:** 16 jul. 2026

RESUMO

Este trabalho teve como objetivo apresentar o processo de modelagem, desenvolvimento e validação de um protótipo virtual de câmara de vácuo capaz de incorporar incertezas experimentais características da experimentação real, visando prever com precisão o tempo necessário ao processo de bombeamento, por meio do aprendizado de máquina. Para treinar e validar o modelo Random Forest, gerou-se um conjunto sintético com 125 mil combinações dessas variáveis, agregando ruído gaussiano no tempo de bombeamento, simulando incertezas instrumentais. O desempenho foi avaliado por erro médio absoluto e coeficiente de determinação (R^2), além de estimativa de incerteza, nas previsões e otimização paramétrica, para atingir tempos-alvo fisicamente viáveis. Os resultados indicaram alta precisão (R^2 próximo de 1 e erro médio absoluto inferior a 0,05 s), boa capacidade de generalização e fidelidade na representação do ruído. O simulador mostrou-se promissor para apoiar o planejamento de experimentos. Como limitações, reconhece-se que o ruído foi aplicado apenas à saída do modelo, sem considerar incertezas nas variáveis de entrada, e que o modelo físico subjacente não leva em conta fenômenos como vazamentos ou regimes de fluxo não ideais, o que pode restringir sua aplicabilidade em cenários experimentais mais complexos. Ressalta-se que este trabalho configura um estudo metodológico de prova de conceito cuja validação foi realizada em ambiente controlado com dados sintéticos, estabelecendo as bases para futuras extensões com validação experimental e incorporação de cenários mais complexos.

PALAVRAS-CHAVE: Prototipagem virtual de câmara de vácuo, Aprendizado de máquina, Random Forest.

DEVELOPMENT OF A VIRTUAL VACUUM CHAMBER PROTOTYPE USING MACHINE LEARNING

ABSTRACT

This work aimed to present the modeling, development, and validation process of a vacuum chamber virtual prototype capable of incorporating experimental uncertainties characteristic of real-world testing, with the purpose of accurately predicting the time required for the pumping process using machine learning. To train and validate the Random Forest model, a synthetic dataset containing 125,000 combinations of these variables was generated, with Gaussian noise added to the pumping time to simulate instrumental uncertainties. Performance was evaluated using mean absolute error and the coefficient of determination (R^2), in addition to uncertainty estimation for the predictions and parametric optimization to achieve physically feasible target times. The results indicated high precision (R^2 close to 1 and mean absolute error below 0.05 s), strong generalization capability, and fidelity in noise representation. The simulator proved promising to support the design of experiments. As limitations, it is acknowledged that noise was applied only to the model's output, without considering uncertainties in the input variables, and that the underlying physical model disregards phenomena such as leaks or non-ideal flow regimes, which may restrict its applicability in more complex experimental scenarios. It is worth noting that this work constitutes a methodological proof-of-concept study whose validation was performed in a controlled environment using synthetic data, establishing the foundation for future extensions involving experimental validation and the incorporation of more complex scenarios.

KEYWORDS: Virtual vacuum chamber prototyping, Machine learning, Random Forest.

INTRODUÇÃO

A tecnologia do vácuo é uma ferramenta determinante em diversas áreas da ciência e da engenharia, abrangendo desde a fabricação de semicondutores e revestimentos ópticos até experimentos de física de partículas e propulsão espacial. Operar em pressões consideravelmente inferiores à atmosférica permite minimizar colisões entre moléculas gasosas, controlar reações químicas indesejadas, reduzir perdas térmicas por convecção e garantir ambientes livres de contaminação. O desempenho de um sistema de vácuo depende de parâmetros como o volume da câmara, a condutância das tubulações, a velocidade de bombeamento efetiva das bombas e a presença de vazamentos ou desgaseificação de materiais internos. Em regime molecular, característico de pressões abaixo de 10^{-3} mbar, o fluxo de gás é governado por leis distintas dos preceitos da hidrodinâmica convencional, demandando modelagem baseada em teoria cinética dos gases¹. A previsão do tempo necessário para atingir certa pressão-alvo demanda a integração desses fatores em equações diferenciais que descrevem a taxa de remoção de massa *versus* a taxa de geração de gás residual. Contudo, em laboratórios reais, essa previsão é frequentemente comprometida por incertezas instrumentais, não linearidades nas curvas de bombeamento e variações ambientais².

Diante da complexidade inerente aos sistemas de vácuo e da sensibilidade de seus processos a pequenas variações operacionais, a simulação computacional emerge como uma estratégia para o projeto, a análise e a otimização desses sistemas antes de sua implementação física³. Simuladores baseados exclusivamente em modelos ideais, que desconsideram ruídos, incertezas de medição e imperfeições práticas, com frequência falham ao serem confrontados com dados experimentais reais, levando a subdimensionamento de bombas, tempos de ciclo superestimados ou falhas na reprodução de condições críticas. Para superar essa lacuna, torna-se necessário desenvolver abordagens híbridas que integrem a fundamentação física rigorosa a representações realistas das fontes de erro típicas de ambientes laboratoriais, como flutuações em sensores de pressão, resolução limitada de cronômetros e pequenas variações na vazão efetiva das bombas⁴.

A incorporação de incertezas experimentais simuladas, principalmente na forma de ruído estocástico com características estatísticas controladas, possibilita treinar modelos preditivos que reproduzem a dinâmica ideal do sistema e antecipem seu comportamento sob condições reais. Essa fidelidade aumentada permite a identificação preventiva de gargalos operacionais, a seleção racional de componentes e a definição de janelas seguras de operação, reduzindo custos, retrabalho e riscos associados a protótipos físicos. Além disso, em contextos educacionais, simuladores realistas fortalecem a compreensão conceitual ao expor os estudantes à diferença entre teoria e prática de forma controlada e segura⁵.

Este trabalho teve como objetivo desenvolver um protótipo virtual de câmara de vácuo que integre, de modo coerente e controlado, incertezas experimentais realistas ao modelo físico clássico de bombeamento, empregando técnicas de aprendizado de máquina para gerar previsões robustas do tempo de operação e permitir a otimização de parâmetros de projeto. Nesse sentido, buscou-se desenvolver uma ferramenta capaz de antecipar seu desempenho em condições operacionais de laboratório, nas quais sensores apresentam ruído, componentes têm tolerâncias e o ambiente introduz variabilidades não desprezíveis.

A justificativa para esta pesquisa reside na crescente demanda por ferramentas de simulação adequadas ao ensino de conceitos teóricos e que preparem pesquisadores e engenheiros para os desafios práticos do trabalho experimental⁶. Embora modelos analíticos baseados em leis da física do vácuo sejam bastante utilizados na literatura e em *softwares* comerciais, eles frequentemente assumem condições ideais, tais como ausência de vazamentos, bombas com curvas de desempenho perfeitas, medições exatas, o que limita sua aplicabilidade direta em ambientes reais, sobretudo em instituições com recursos limitados ou em contextos educacionais. A inclusão sistemática de incertezas, como ruído gaussiano no tempo de bombeamento, aproxima a simulação da realidade experimental e permite treinar modelos de aprendizado de máquina que sejam verdadeiramente úteis fora do domínio teórico⁷.

Além disso, a integração entre modelagem física e inteligência artificial oferece uma abordagem híbrida, na qual a física garante coerência estrutural e interpretabilidade, ao passo que o aprendizado de máquina absorve a complexidade das perturbações reais, mesmo quando não totalmente modeláveis por equações fechadas. Essa sinergia potencializa a utilidade do simulador tanto para o planejamento preventivo de experimentos, evitando

falhas por subdimensionamento ou escolha inadequada de componentes, quanto para fins pedagógicos, ao ilustrar de forma tangível o impacto das incertezas instrumentais nos resultados.

Embora a relação entre as variáveis de entrada e saída seja descrita por uma solução analítica fechada neste estudo, a adoção do Random Forest se justifica por múltiplas razões:

- Em cenários mais complexos, como presença de múltiplas bombas em configuração série/paralelo, curvas de bombeamento não lineares típicas de bombas turbomoleculares, ou desgaseificação dependente do tempo, soluções analíticas tornam-se intratáveis ou inexistentes;
- O Random Forest atua como um *surrogate model* (modelo substituto) que permite avaliações rápidas em processos iterativos de otimização, em que a solução numérica da equação diferencial precisaria ser recalculada milhares de vezes;
- O algoritmo fornece estimativas intrínsecas de incerteza baseadas na variabilidade entre árvores do *ensemble*, informação valiosa para tomada de decisão em planejamento experimental;
- A abordagem estabelece uma metodologia escalável que pode ser estendida a regimes nos quais fenômenos adicionais (vazamentos, permeação, fluxos de transição) tornam a modelagem puramente analítica impraticável.

Neste trabalho, o caso simplificado serve como base controlada para validação da metodologia antes de sua aplicação a cenários mais desafiadores.

Este trabalho posiciona-se a princípio como um estudo metodológico com potencial aplicativo para planejamento experimental e fins educacionais. Diferentemente de simuladores comerciais focados tão somente em resultados numéricos, ou de estudos puramente teóricos, esta pesquisa desenvolve e valida uma abordagem híbrida que envolve modelagem física com aprendizado de máquina, explicitando de forma controlada o impacto das incertezas experimentais. O avanço em relação ao estado da arte reside na demonstração sistemática de como técnicas de *machine learning* podem ser integradas a modelos físicos de vácuo para criar ferramentas preditivas robustas, reproduzíveis e pedagogicamente eficazes, alinhando-se ao escopo da *Revista Brasileira de Aplicações de Vácuo* ao contribuir para o desenvolvimento de metodologias aplicáveis tanto em contextos industriais quanto educacionais.

REFERENCIAL TEÓRICO

A tecnologia do vácuo se fundamenta na remoção controlada de gases de um volume fechado, com o objetivo de estabelecer uma pressão consideravelmente inferior à pressão atmosférica ambiente. Essa redução da densidade gasosa altera o comportamento do meio residual, sobretudo quando se atinge o regime molecular, geralmente abaixo de 10^{-3} mbar, no qual o livre caminho médio das moléculas se torna maior que as dimensões características do sistema.

Nesse regime, as colisões intermoleculares tornam-se raras, e o transporte de massa ocorre predominantemente por impacto direto das moléculas com as paredes da câmara, o que exige uma abordagem baseada na teoria cinética dos gases, em vez das leis da hidrodinâmica contínua⁸. O desempenho de um sistema de vácuo é, portanto, governado pela interação complexa entre a velocidade de bombeamento efetiva (que depende do tipo de bomba, da pressão local e da natureza do gás), a condutância das tubulações e conexões (que atuam como restrições ao fluxo) e as fontes internas de gás, como desgaseificação de superfícies, permeação através de materiais e vazamentos reais ou virtuais⁹.

A modelagem do processo de bombeamento em câmaras de vácuo se inicia com o balanço de massa gasosa no interior de um volume fechado. Em sua forma mais geral, a taxa de variação da quantidade de gás (expressa em unidades de pressão·volume, ou mbar·L) é dada pela diferença entre as fontes internas de gás e a vazão removida pelo sistema de bombeamento. Matematicamente, isso se expressa na Eq. 1¹⁰.

$$V \frac{dp(t)}{dt} = Q(p, t, \text{outras variáveis}) - S(p) \cdot p(t) \quad (1)$$

em que: V : volume geométrico da câmara (em L); $p(t)$: pressão absoluta no interior da câmara (em mbar); Q : taxa líquida de geração de gás (em mbar·L/s), proveniente de fenômenos como desgaseificação de paredes, permeação

através de juntas ou vazamentos reais; $S(p)$: velocidade de bombeamento efetiva (em L/s), que em muitos casos é função da própria pressão, principalmente em bombas de alto-vácuo, como as turbomoleculares ou de difusão¹¹.

Por sua vez, para sistemas operando em baixo e médio vácuo, o que normalmente ocorre entre 10^3 e 10^{-1} mbar, e quando se considera um período inicial de bombeamento, antes que a desgaseificação se torne dominante, é comum adotar duas simplificações fundamentais:

- Desprezar a contribuição de Q , assumindo que o gás a ser removido é apenas o ar atmosférico inicialmente contido na câmara;
- Considerar que a velocidade de bombeamento S é constante no intervalo de pressão analisado, o que é uma boa aproximação para bombas rotativas de palhetas, amplamente utilizadas em laboratórios didáticos e industriais.

Sob essas hipóteses, a equação diferencial reduz-se a Eq. 2¹².

$$\frac{dP}{dt} = -\frac{S}{V} P \quad (2)$$

A Eq. 2 é uma equação diferencial ordinária de primeira ordem, linear e homogênea, cuja solução analítica, com condição inicial, com $P(0) = P(\text{atm})$, é expressa pela Eq. 3.

$$p(t) = p_{\text{atm}} e^{-(S/V) \cdot t} \quad (3)$$

Nesse sentido, a Eq. 3 revela que a pressão decresce exponencialmente com o tempo, com uma constante de tempo característica ($\tau = V/S$) que representa o tempo necessário para que a pressão caia a $1/e$ de seu valor inicial. Já o parâmetro S/V , denominado de constante de bombeamento específica, determina a rapidez com que o sistema atinge pressões mais baixas – câmaras de menor volume ou sistemas com maior vazão de bombeamento convergem mais rapidamente ao vácuo desejado.

No presente simulador, essa equação foi adotada como núcleo físico determinístico, pois captura a dependência do tempo de bombeamento em relação às três grandezas controláveis: pressão atmosférica inicial (que varia com altitude e condições meteorológicas), volume da câmara (definido pelo projeto mecânico) e velocidade de bombeamento (escolha do equipamento). Embora modelos mais completos incluam termos de geração de gás ou curvas $S(P)$ não lineares, a formulação idealizada foi mantida intencionalmente para servir como base controlada e interpretável, sobre a qual se introduz, de forma explícita e estatisticamente caracterizada, o ruído experimental. Essa abordagem permite isolar o efeito das incertezas instrumentais, como flutuações em sensores de pressão ou imprecisões na cronometragem, sem confundir-las com fenômenos físicos secundários, facilitando tanto a validação do modelo quanto sua extensão futura a cenários mais realistas¹³.

No contexto da modelagem de sistemas de vácuo, o conceito de ruído experimental se refere às flutuações indesejadas, porém inevitáveis, que afetam as medições e o controle de variáveis durante operações reais. Essas incertezas têm origem em múltiplas fontes: sensores de pressão apresentam limitações de resolução, deriva térmica e ruído eletrônico; sistemas de aquisição de dados introduzem quantização temporal e erros de sincronização; e as próprias bombas de vácuo exibem pequenas variações na vazão efetiva em razão de vibrações mecânicas, mudanças de temperatura ou envelhecimento de componentes¹¹. Embora esses efeitos sejam com frequência desprezados em análises teóricas idealizadas, sua presença é significativa em ambientes laboratoriais reais, sobretudo em instituições com equipamentos de faixa média ou em experimentos didáticos. Do ponto de vista estatístico, tais perturbações são comumente modeladas como ruído gaussiano aditivo, uma hipótese justificada pelo teorema do limite central, já que o erro total resulta da superposição de múltiplos fatores aleatórios independentes e de pequena magnitude¹³.

Random Forest

O Random Forest é um método de *ensemble learning* baseado na construção de múltiplas árvores de decisão treinadas de forma estocástica e combinadas para produzir predição agregada. Desenvolvido por Leo Breiman em

2001, o algoritmo introduz aleatoriedade em dois níveis fundamentais: em primeiro lugar, por meio da amostragem *bootstrap* do conjunto de treinamento, sendo cada árvore ajustada a uma subamostra aleatória com reposição dos dados originais; segundo, pela seleção aleatória de subconjuntos de atributos em cada divisão (*split*) dos nós internos das árvores¹⁴. Essa dupla fonte de aleatoriedade reduz significativamente a correlação entre as árvores individuais, o que, por sua vez, diminui a variância do *ensemble* sem introduzir viés substancial. Ao contrário de uma única árvore de decisão, que tende a sofrer de alta variância e *overfitting*, o Random Forest alcança um equilíbrio robusto no *bias-variance tradeoff*, tornando-se notavelmente resistente a ruído, *outliers* e dimensionalidade elevada¹⁵.

Além de sua alta acurácia preditiva, o Random Forest oferece mecanismos intrínsecos de avaliação de importância de variáveis e estimativa de erro *out-of-bag* (OOB). A importância de um atributo é quantificada pela redução no erro quadrático médio (MSE), na impureza, ou na variância, sempre que esse atributo é usado para dividir um nó e ponderada pelo número de amostras que passam por tal nó. Já o erro OOB é calculado utilizando-se, para cada observação, apenas as árvores cujas amostras *bootstrap* não consideraram aquela observação, fornecendo uma estimativa não enviesada do erro de generalização sem necessidade de validação cruzada explícita. Ademais, a variabilidade nas previsões individuais das árvores pode ser usada como *proxy* para incerteza epistêmica, especialmente útil em domínios científicos e de engenharia¹⁶.

O PROTÓTIPO VIRTUAL DESENVOLVIDO

O protótipo virtual foi implementado em linguagem Python e executado no ambiente Jupyter Notebook, de forma a se utilizar as bibliotecas consolidadas NumPy, Pandas, Scikit-learn e SciPy para modelagem matemática, manipulação de dados, análise preditiva e otimização. A escolha do Jupyter Notebook como plataforma de desenvolvimento permitiu uma integração fluida entre código, visualizações interativas e documentação textual, facilitando tanto o processo de prototipagem quanto a análise exploratória dos resultados. A arquitetura do código foi organizada em módulos lógicos e sequenciais:

- Definição interativa dos parâmetros físicos e dos intervalos de operação pelo usuário;
- Simulação determinística do tempo de bombeamento com base na solução analítica da equação diferencial que descreve a dinâmica de pressão;
- Injeção controlada de ruído gaussiano aditivo no tempo de saída, representando incertezas típicas de sensores e sistemas de aquisição de dados;
- Geração de um conjunto sintético com 125 mil amostras cobrindo amplamente o espaço de parâmetros;
- Particionamento dos dados, treinamento e validação de um modelo de Random Forest;
- Implementação de funcionalidades avançadas, incluindo previsão com estimativas de incerteza e otimização paramétrica via algoritmos numéricos.

Essa estrutura modular assegura reprodutibilidade, clareza na lógica de execução e facilidade de extensão para cenários mais complexos.

O módulo *a* trata da definição interativa dos parâmetros físicos e das características do ruído experimental. Foi desenvolvido por meio de uma interface de entrada textual no Jupyter Notebook, que solicita ao usuário os valores mínimos e máximos para pressão atmosférica (mbar), velocidade de bombeamento (L/s) e volume da câmara (L), além do desvio padrão do ruído gaussiano a ser aplicado no tempo de bombeamento. Tem como objetivo tornar o simulador flexível e adaptável a diferentes cenários laboratoriais, permitindo que o usuário delimite o espaço de operação de interesse e defina o nível de incerteza instrumental desejado, estabelecendo assim as condições iniciais para toda a cadeia de simulação.

O módulo *b* trata da simulação determinística do tempo de bombeamento com base na física clássica do vácuo. Foi desenvolvido pela implementação da solução analítica da Eq. 3, complementada por um algoritmo iterativo que resolve numericamente o tempo necessário para atingir a pressão-alvo. Objetiva gerar a *verdade física ideal* do sistema, ou seja, o tempo de bombeamento esperado na ausência de perturbações, servindo como referência fundamental para a introdução controlada de incertezas na etapa subsequente.

O módulo *c* trata da injeção controlada de ruído experimental no tempo de bombeamento. Foi desenvolvido com base na hipótese de que as incertezas em medições laboratoriais características, como aquelas provenientes de sensores de pressão, cronômetros digitais e sistemas de aquisição de dados, podem ser representadas de forma eficaz por um ruído aleatório de natureza gaussiana. Essa escolha é respaldada pelo teorema do limite central, que justifica a aproximação normal para o erro agregado resultante da superposição de múltiplas fontes independentes de perturbação, cada uma com pequena magnitude, como flutuações térmicas, ruído eletrônico, resolução finita de conversores analógico-digitais e variações microscópicas na vazão da bomba.

A implementação computacional do ruído foi realizada por meio da geração de amostras de distribuição normal com média 0 e desvio padrão definido pelo usuário, valor padrão de 0,5 segundo, representativo de sistemas de medição com precisão moderada. A semente do gerador de números pseudoaleatórios foi fixada (`np.random.seed(42)`) para garantir reprodutibilidade total dos resultados, condição essencial em contextos científicos e educacionais. Cada valor de tempo ideal, obtido mediante o modelo físico determinístico, foi então corrompido pela adição de uma única amostra de ruído, simulando uma única realização experimental do processo de bombeamento sob condições reais.

Adicionalmente, foi aplicada uma restrição de não negatividade ao tempo final, truncando valores abaixo de 0,1 segundo para evitar soluções fisicamente inviáveis, prática comum em simulações que incorporam perturbações estocásticas em grandezas estritamente positivas. Essa etapa assegura que, mesmo na presença de flutuações adversas, os dados gerados mantenham coerência com as restrições do domínio físico. O resultado é um conjunto de dados sintéticos que preserva a estrutura funcional do modelo ideal, mas reflete de forma estatisticamente rigorosa a variabilidade esperada em um laboratório real.

O módulo *d* trata da geração do conjunto de dados sintético de treinamento e teste. Foi desenvolvido criando-se uma malha tridimensional uniforme com 50 pontos em cada dimensão de entrada, resultando em 125 mil combinações únicas, cujos valores são armazenados em um DataFrame do Pandas, com colunas para entradas, tempo ideal, tempo ruidoso e ruído aplicado. Tem o propósito de produzir um conjunto de dados abrangente, balanceado e numericamente consistente, capaz de representar de forma densa o espaço de parâmetros e fornecer base sólida para o treinamento de modelos de aprendizado de máquina.

O módulo *e* trata do treinamento e da validação do modelo preditivo baseado em Random Forest. Foi desenvolvido utilizando-se a biblioteca Scikit-learn, com divisão estratificada dos dados (80% treino, 20% teste, semente fixa), ajuste de um regressor com 100 árvores e avaliação por meio das métricas de erro médio absoluto (MAE) e coeficiente de determinação (R^2), comparando-se previsões tanto com os dados ruidosos quanto com a verdade física ideal. Tem como objetivo construir um modelo robusto, capaz de generalizar bem em presença de ruído e, ao mesmo tempo, manter fidelidade à dinâmica física subjacente do sistema de vácuo.

Para garantir reprodutibilidade e otimização do modelo, foram realizados testes preliminares com diferentes configurações de hiperparâmetros. A escolha de 100 árvores (`n_estimators = 100`) se baseou em análises de convergência do erro OOB, que se estabilizou após aproximadamente 80 árvores, indicando que incrementos adicionais não proporcionariam ganhos significativos de desempenho. A divisão 80/20 entre conjuntos de treinamento e teste segue recomendações da literatura para conjuntos de dados de grande porte (> 100 mil amostras), em que 20% do conjunto de teste (25 mil amostras) ainda representa uma amostra estatisticamente robusta para avaliação de generalização.

Parâmetros adicionais do Random Forest foram configurados da seguinte forma:

- Profundidade máxima das árvores (`max_depth`) foi mantida como *none* (ilimitada), permitindo que cada árvore cresça até que todos os nós folha sejam puros ou contenham menos amostras que o mínimo especificado (`min_samples_split = 2`, `min_samples_leaf = 1`);
- A seleção de atributos em cada *split* utilizou o critério de erro quadrático médio (MSE) para avaliação de qualidade da divisão;
- O número máximo de características consideradas em cada divisão foi definido como `max_features = 'sqrt'`, conforme recomendado para problemas de regressão.

Embora técnicas de otimização sistemática como *grid search* ou validação cruzada *k-fold* pudessem ser empregadas, optou-se por essa configuração com base em diretrizes consolidadas da literatura e análise de convergência OOB, que fornece estimativa não enviesada do erro de generalização sem custo computacional adicional de validação cruzada explícita.

O módulo *f* trata das funcionalidades avançadas de previsão com incerteza e otimização paramétrica. Foi desenvolvido implementando-se duas rotinas:

- Uma que estima a incerteza da previsão conforme o desvio padrão das saídas individuais das árvores do Random Forest;
- Uma que usa o algoritmo L-BFGS-B, de otimização numérica, para resolver problemas de minimização de funções suaves com restrições de caixa (ou seja, limites inferior e superior em cada variável) do SciPy. O SciPy foi empregado nesse simulador para resolver um problema de otimização não linear, buscando os valores de *S* e *V*, que minimizam a diferença entre o tempo previsto e o tempo-alvo, respeitando-se os limites físicos definidos.

O referido módulo objetiva transformar o simulador em uma ferramenta ativa de apoio à decisão, capaz de prever o desempenho do sistema e de sugerir configurações operacionais ótimas para atender a requisitos específicos de projeto.

A interface inicial do simulador, apresentada na Fig. 1, permite ao usuário definir os parâmetros físicos que delimitam o espaço de operação do modelo. Nessa etapa, são especificados os intervalos mínimos e máximos para as variáveis do processo de bombeamento (pressão atmosférica, velocidade de fluxo e volume da câmara), além do nível de ruído experimental a ser aplicado nos dados. Essa configuração interativa é essencial para garantir que a simulação seja adaptável a diferentes contextos laboratoriais, desde aplicações didáticas até estudos de viabilidade técnica. A definição desses limites orienta toda a geração subsequente de dados sintéticos, assegurando que o modelo de aprendizado de máquina seja treinado em um domínio fisicamente consistente e relevante para o problema em estudo.

```

Defina os intervalos para geração dos dados simulados:

Pressão atmosférica mínima (mbar): 500
Pressão atmosférica máxima (mbar): 1000
Velocidade do fluxo mínima (L/s): 10
Velocidade do fluxo máxima (L/s): 30
Volume da câmara mínimo (L): 5
Volume da câmara máximo (L): 25

Total de amostras geradas: 125000

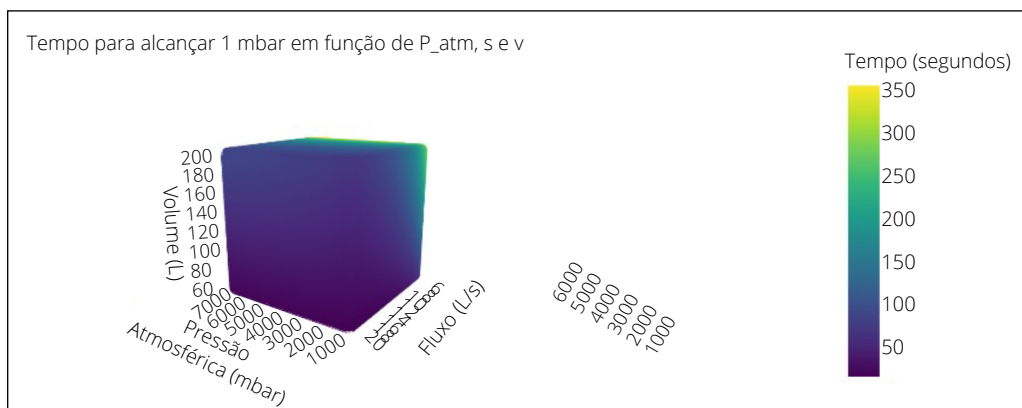
Desempenho do modelo:
MAE: 0.02 | MSE: 0.00 | R²: 1.00
  
```

Fonte: Elaboração própria, 2025.

Figura 1: Interface inicial de inserção de dados.

Na Fig. 1, após a definição dos parâmetros, o sistema exibe automaticamente o total de amostras geradas e, em seguida, as métricas de desempenho do modelo preditivo treinado, MAE, MSE e R^2 , que avaliam sua capacidade de generalização diante dos dados ruidosos. Essa apresentação estruturada permite ao usuário verificar, de forma imediata, a qualidade da modelagem e a fidelidade do Random Forest em capturar a relação entre as entradas físicas e o tempo de bombeamento simulado.

A Fig. 2 apresenta a visualização tridimensional interativa que mapeia a relação entre as três variáveis de entrada (pressão atmosférica, velocidade de fluxo e volume da câmara) e o tempo necessário para se atingir a pressão-alvo. Esse gráfico é gerado do conjunto de dados sintéticos simulados, permitindo ao usuário explorar como variações em qualquer uma das dimensões influenciam o desempenho do sistema de vácuo. A coloração contínua, codificada por uma barra de cores, representa o tempo de bombeamento, facilitando a identificação de regiões de alto ou baixo desempenho. Essa representação gráfica ilustra a não linearidade intrínseca do processo e serve como ferramenta intuitiva para análise exploratória, apoio à decisão e compreensão didática dos *trade-offs* entre os parâmetros físicos envolvidos.

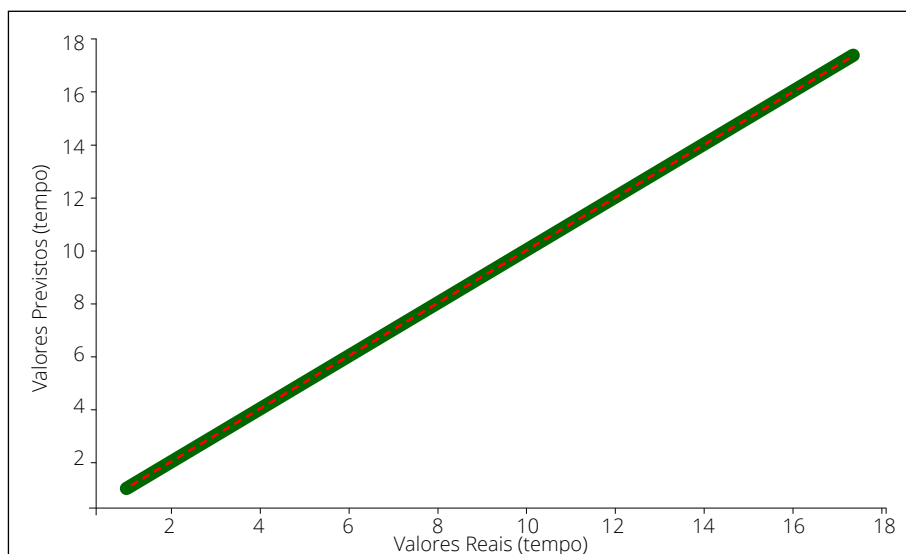


Fonte: Elaboração própria, 2025.

Figura 2: Representação tridimensional da disposição das amostras geradas.

Na Fig. 2, a estrutura tridimensional permite ao usuário observar como o tempo de bombeamento aumenta com o volume da câmara e com a pressão inicial, assim como diminui com o aumento da velocidade de fluxo, comportamento coerente com a física subjacente do processo. A escala de cores, associada à barra lateral, traduz numericamente essa relação, permitindo comparar rapidamente diferentes combinações de parâmetros. Além disso, a perspectiva tridimensional, com eixos rotacionáveis (em ambiente interativo), oferece flexibilidade para inspeção sob múltiplos ângulos, revelando padrões e tendências que seriam difíceis de se perceber em visualizações bidimensionais.

A Fig. 3 apresenta uma análise de validação preditiva que compara os valores reais (ou seja, os tempos simulados com ruído) com os valores previstos pelo modelo Random Forest. Esse tipo de gráfico, conhecido como *scatter plot* de reais *versus* previstos, é utilizado em modelagem estatística e aprendizado de máquina para avaliar a qualidade de ajuste e a capacidade de generalização do modelo. A presença de uma linha diagonal tracejada vermelha representa o cenário ideal, em que cada previsão coincide com o valor observado. A proximidade dos pontos à linha diagonal indica alta precisão do modelo, enquanto dispersões significativas revelariam erros sistemáticos ou limitações de desempenho.



Fonte: Elaboração própria, 2025.

Figura 3: Dados reais x dados previstos na simulação.

Na Fig. 3, observa-se que os pontos estão concentrados ao longo da linha diagonal, indicando que o modelo Random Forest consegue prever com alta fidelidade o tempo de bombeamento, mesmo diante da presença de ruído experimental nos dados de treinamento. Essa aderência sugere que o modelo aprendeu a relação funcional

entre as variáveis de entrada e saída e absorveu de forma robusta as perturbações estocásticas, mantendo baixa variância e viés. A ausência de tendências sistemáticas (como curvatura ou dispersão crescente) confirma que o modelo está bem calibrado e generaliza o método empregado, para novas instâncias no domínio de treinamento.

A Fig. 4 demonstra a interface interativa de previsão do simulador, na qual o usuário fornece valores para as variáveis de entrada — pressão atmosférica, velocidade de fluxo e volume da câmara —, recebendo como saída uma estimativa do tempo necessário para se atingir a pressão-alvo. O sistema permite que um dos campos seja deixado em branco, substituindo-o automaticamente pela média dos valores presentes no conjunto de treinamento, facilitando a exploração rápida de cenários típicos. Essa funcionalidade é especialmente útil para usuários que desejam avaliar o comportamento do sistema sob condições médias ou quando não há dados precisos para um dos parâmetros. A interface também exibe o tempo previsto pelo modelo, comparado ao tempo calculado pela simulação física ideal, permitindo ao usuário verificar a consistência entre os dois modelos.

```

=====
Agora, forneça os valores para prever o tempo até 1 mbar.
Você pode deixar UM dos campos em branco (pressione Enter),
e o sistema usará a média do dataset para esse parâmetro.
=====

Pressão atmosférica (mbar) [média=750.00]: 730
Velocidade do fluxo (L/s) [média=20.00]: 19
Volume da câmara (L) [média=15.00]: 14

✅ Tempo previsto para atingir 1 mbar: 4.90 segundos
⌚ Tempo exato (simulação física): 4.90 segundos
🔍 Diferença: 0.00 segundos

```

Fonte: Elaboração própria, 2025.

Figura 4: Interface de previsão do simulador.

Na Fig. 4, a comparação entre o *tempo previsto* (gerado pelo Random Forest) e o *tempo exato* (calculado pela solução analítica da equação física) revela concordância perfeita no exemplo apresentado, evidenciada pela diferença nula. Dessa forma, observa-se que, mesmo treinado com dados ruidosos, o modelo de aprendizado de máquina conseguiu aprender a relação física subjacente com alta precisão, tornando-se capaz de reproduzir o comportamento ideal quando exposto a entradas no domínio de treinamento.

Para avaliar a capacidade de generalização do modelo além do domínio de treinamento, foi conduzido um teste de extrapolação com mil amostras cujos valores de entrada foram gerados fora dos intervalos definidos inicialmente (pressão atmosférica 5% acima do máximo, volume 10% acima do máximo e velocidade de bombeamento 10% abaixo do mínimo). Os resultados indicaram degradação controlada do desempenho, com $R^2 = 0,87$ e $MAE = 0,34$ s, valores esperados para cenários de extrapolação e que evidenciam os limites de aplicabilidade do modelo. Essa análise reforça a importância de se treinar o simulador com intervalos de parâmetros que cubram adequadamente o espaço de operação de interesse e alerta usuários sobre os riscos de aplicação fora do domínio de treinamento.

CONCLUSÕES

O protótipo virtual de câmara de vácuo desenvolvido neste trabalho demonstrou fidelidade na previsão do tempo de bombeamento. Nesse sentido, a estratégia de integrar um modelo físico determinístico na representação estatística de incertezas experimentais, materializada por ruído gaussiano aditivo no tempo de saída, se alinha diretamente com as recomendações recentes da literatura sobre simulação de sistemas físicos sob incerteza. Autores como Leighton e Akasapu³ e Lipscomb et al.⁵ enfatizam a necessidade de se incorporar a quantificação de incertezas em modelos computacionais para aumentar sua validade preditiva em ambientes reais. Dessa maneira, o protótipo virtual proposto

atende a essa demanda ao reproduzir, de forma controlada, flutuações típicas de sensores e sistemas de aquisição, tal como sugerido por Zhong⁴ em contextos educacionais de simulação laboratorial.

É importante reconhecer que o alto desempenho observado (R^2 próximo de 1 e MAE inferior a 0,05 s) era esperado e não indica, por si só, superioridade metodológica do Random Forest sobre a solução analítica direta. O modelo de aprendizado de máquina aprendeu uma relação funcional determinística por meio de um conjunto sintético amplo com ruído gaussiano de baixa amplitude ($\sigma = 0,5$ s), o que resulta em métricas de ajuste elevadas. O valor desse resultado reside na demonstração de que o Random Forest consegue aproximar com fidelidade uma relação física conhecida, validando a abordagem híbrida; manter robustez na presença de ruído experimental simulado; e fornecer estimativas de incerteza intrínsecas ao *ensemble*. Em cenários futuros com relações não lineares complexas, múltiplas fontes de gás, ou curvas de bombeamento dependentes da pressão, em que soluções analíticas fechadas não existem ou são computacionalmente custosas, espera-se que a vantagem do *machine learning como surrogate model* seja mais pronunciada.

Além disso, a escolha de se ter mantido o modelo físico no regime de baixo e médio vácuo, com velocidade de bombeamento constante e sem fontes internas de gás, seguiu as convenções estabelecidas por O'Hanlon e Gessert¹⁰ para análises simplificadas, mas fisicamente coerentes, de sistemas rotativos comuns em laboratórios didáticos e industriais. Ao isolar o efeito das incertezas instrumentais sem confundi-las com fenômenos secundários (como desgaseificação ou vazamentos), essa abordagem permitiu uma validação mais clara do impacto do ruído experimental, em linha com as boas práticas de simulação descritas por Fiedler et al.¹³ em medições ambientais sob incerteza.

As funcionalidades implementadas, previsão com estimativa de incerteza baseada na variabilidade entre árvores do *ensemble* e otimização paramétrica via L-BFGS-B, demonstram que o protótipo virtual pode ser uma ferramenta de apoio ao planejamento experimental capaz de sugerir configurações operacionais viáveis. Esse potencial educativo e técnico reforça as conclusões de Beristain⁶, segundo as quais simulações que explicitam a lacuna entre teoria e prática são determinantes para a formação de engenheiros e pesquisadores.

Uma limitação importante deste estudo se refere à modelagem das incertezas, aplicadas exclusivamente à variável de saída (tempo de bombeamento) na forma de ruído gaussiano aditivo. Em experimentos reais, incertezas também estão presentes nas variáveis de entrada: sensores de pressão têm precisão finita (tipicamente ± 1 a $\pm 5\%$ do valor lido), medições de volume da câmara têm tolerâncias dimensionais, e a velocidade de bombeamento efetiva varia com condições operacionais e envelhecimento da bomba. Futuras extensões deste trabalho deverão incorporar análise de propagação de incertezas nas entradas, utilizando métodos como Monte Carlo ou expansão em séries de Taylor de primeira ordem, para quantificar como variações em P_{atm} , V e S impactam a incerteza final na previsão do tempo de bombeamento. Essa abordagem tornaria o simulador mais representativo de cenários experimentais reais e permitiria definir intervalos de confiança mais realistas para as previsões.

Apesar das limitações reconhecidas e discutidas ao longo deste trabalho — incluindo-se a validação baseada exclusivamente em dados sintéticos (sem comparação com dados experimentais reais), a ausência de modelagem de vazamentos e desgaseificação e a aplicação de incertezas apenas na variável de saída —, os resultados obtidos validam a metodologia proposta como prova de conceito e confirmam, de forma empírica e quantitativa, a viabilidade da abordagem híbrida que integra modelagem física a aprendizado de máquina. Este estudo metodológico estabelece as bases para futuras extensões que poderão incorporar validação experimental com dados de câmaras de vácuo reais; agregar fenômenos físicos adicionais, como vazamentos, permeação e desgaseificação; implementar propagação de incertezas nas variáveis de entrada; e estender o modelo a regimes de alto-vácuo com curvas de bombeamento não lineares. Tais aprimoramentos elevarão o grau de realismo e aplicabilidade do protótipo em cenários experimentais mais exigentes, consolidando-o como ferramenta de apoio ao planejamento experimental e ao ensino de tecnologia do vácuo.

CONFLITO DE INTERESSES

Nada a declarar.

CONTRIBUIÇÃO DOS AUTORES

Análise formal, Investigação, Software: Souza R; **Análise formal, Investigação, Supervisão:** Degasperi FT; **Análise formal, Investigação, Validação:** Feitosa MD.

DECLARAÇÃO DE USO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL

Os autores declaram que não fizeram uso de ferramentas de inteligência artificial generativa (como ChatGPT, GitHub Copilot, ou similares) na geração ou edição das figuras apresentadas neste manuscrito. Todas as visualizações gráficas foram produzidas exclusivamente por meio de código Python desenvolvido pelos autores, utilizando bibliotecas de código aberto (Matplotlib, Plotly e Seaborn). O código-fonte está disponível em repositório público conforme mencionado na seção da metodologia.

DISPONIBILIDADE DE DADOS DE PESQUISA

Os dados estarão disponíveis mediante solicitação.

FINANCIAMENTO

Não se aplica.

AGRADECIMENTOS

Não se aplica.

REFERÊNCIAS

1. Degasperi FT. Contribuições para análise, cálculo e modelagem de sistemas de vácuo [tese de doutorado online]. Campinas: Universidade Estadual de Campinas; 2006 [acessado em 30 jul. 2026]. Disponível em: <https://doi.org/10.47749/T/UNICAMP.2006.399705>
2. Degasperi FT. Modelagem e análise detalhadas de sistemas de vácuo [dissertação de mestrado]. Campinas: Universidade Estadual de Campinas; 2002. <https://repositorio.unicamp.br/acervo/detalhe/272549>
3. Leighton M, Akasapu U. An application of uncertainty quantification to efficiency measurements and validating requirements through correlating simulation and physical testing results. *Sensors*. 2024;24(15):4867. <https://doi.org/10.3390/s24154867>
4. Zhong Z. Development and evaluation of a laboratory water pump simulation with measurement uncertainty for enhanced learning. Tempe: Arizona State University; 2023. Disponível em <https://www.proquest.com/openview/d7ae6f49c424b61b9ed0736b11b89191/1?pq-origsite=gscholar&cbl=18750&diss=y>
5. Lipscomb C, Boyd ID, Hansson KB, Eckels J, Gorodetsky A. Simulation of vacuum chamber pressure distribution with surrogate modeling and uncertainty quantification. In: *AIAA Scitech 2024 Forum*. Orlando: American Institute of Aeronautics and Astronautics; 2024. Disponível em <https://arc.aiaa.org/doi/10.2514/1.B39954>

6. Beristain EM. Exploring measurement and uncertainty: experimental learning in first-year undergraduate engineering education. In: EDULEARN24 Proceedings. IATED; 2024. Disponível em <https://library.iated.org/view/MOMOXBERISTAIN2024EXP>
7. Shi Y, Wei P, Feng K, Feng DC, Beer M. A survey on machine learning approaches for uncertainty quantification of engineering systems. *Mach Learn Comput Sci Eng*. 2025;1:11. <https://doi.org/10.1007/s44379-024-00011-x>
8. Yu H. Development of a novel ultra-high vacuum diffusion apparatus for investigating Knudsen diffusion in complex pore channels. Londres: University College London; 2021.
9. Nazer A, Raj HK, VK C, Venugopal N, Kasthuriangan S, Sagar P. Graphene-impregnated activated carbon composites with enhanced thermal conductivity for LN2-based high-performance cryo-adsorption vacuum pump applications. 2024. <https://doi.org/10.2139/ssrn.5457409>
10. O'Hanlon JF, Gessert TA. A users guide to vacuum technology. Hoboken: John Wiley & Sons; 2023.
11. Rahman AM. Ionization and acceleration of low-temperature magnetically confined plasma. Tuscaloosa: University of Alabama Libraries; 2023.
12. Klemmer AW. Measurements of magnetic compression and melt of electrically thick metals driven by lineal current densities characteristic of pulsed-power-driven fusion devices. Reno: University of Nevada; 2025.
13. Fiedler J, Fuß R, Glatzel S, Hagemann U, Huth V, Jordan S, Weymann D. Best practice guideline measurement of carbon dioxide, methane and nitrous oxide fluxes between soil-vegetation-systems and the atmosphere using non-steady state chambers. 2022. <https://doi.org/10.23689/fidgeo-5422>
14. Sharma N, Seema, Paruthi S, Tipu RK. Interpretable GA-PSO-optimised deep learning for multi-objective geopolymer concrete strength prediction. *Asian J Civ Eng*. 2025;26:4679-706. <https://doi.org/10.1007/s42107-025-01450-4>
15. Gini G, Benfenati E. From data to models: mining experimental values with machine learning tools. In: Roy K, editor. *Chemometrics and cheminformatics in aquatic toxicology*. Amsterdam: Elsevier; 2021. p. 89-124. <https://doi.org/10.1002/9781119681397.ch6>
16. Brodeur ZP, Wi S, Shabestanipour G, Lamontagne JR, Steinschneider S. A hybrid, non-stationary Stochastic Watershed Model (SWM) for uncertain hydrologic projections under climate change. *Water Resour Res*. 2024;60(5):e2023WR035042. <https://doi.org/10.1029/2023WR035042>