Detecção de Falhas no Processo Tennessee Eastman por meio de Resíduos de Modelos NARX/NARMAX Identificados a partir de Dados Abertos

Heitor Gonçalves Santana, Helena Bigliani Ferreira e Gabriel Rios de Souza

***Resumo*—Este trabalho investiga se modelos dinâmicos de entrada e saída identificados apenas com dados de operação normal podem servir de base para detectar e isolar falhas no processo Tennessee Eastman, usando-se o conjunto aberto de Rieth et al., com 52 variáveis amostradas a cada 3 min em 500 simulações normais e 10 000 simulações com 20 tipos de falha. O ciclo de vida dos dados foi conduzido segundo um plano de gestão de dados, com armazenamento num banco orientado a documentos com a API do MongoDB, separação de treino, validação e teste por simulação e publicação dos dados derivados e do código no Zenodo. Para cinco subsistemas do processo, foram identificados modelos ARX, NARX polinomiais e NARMAX pelo algoritmo FROLS, além de um NARX neural, e seus resíduos de predição um passo à frente e de simulação livre alimentaram um índice T^2 filtrado, cujo limiar foi fixado no quantil de 99% das simulações normais de validação. Verificou-se que, em operação normal, nenhum termo polinomial de grau 2 atinge a taxa de redução de erro exigida, de modo que a estrutura NARX se reduz à ARX, e que a rede neural não melhora a predição, o que indica relações praticamente lineares em torno do ponto de operação. Ainda assim, o detector por resíduos alcançou taxa média de detecção de 90,6% nas 17 falhas detectáveis, contra 86,7% da análise de componentes principais (PCA), com taxa de falsos alarmes de 1,3%, e superou a PCA nas falhas que alteram a dinâmica sem deslocar a média (10, 16, 19 e 20); a fusão dos dois detectores atingiu 93,3%, e a contribuição de cada subsistema ao alarme apontou corretamente a origem física de falhas como a 4, a 11 e a 14.**

***Palavras-chave*—Detecção de falhas, identificação de sistemas, NARMAX, FROLS, Tennessee Eastman, análise de resíduos, ciclo de vida dos dados, MongoDB.**

***Abstract*—This work investigates whether dynamic input-output models identified only from normal operating data can serve as the basis for detecting and isolating faults in the Tennessee Eastman process, using the open dataset by Rieth et al., with 52 variables sampled every 3 min in 500 fault-free simulations and 10 000 simulations under 20 fault types. The data life cycle followed a data management plan, with storage in a document-oriented database with the MongoDB API, training, validation and test splits by simulation, and publication of the derived data and code on Zenodo. For five subsystems of the process, ARX, polynomial NARX and NARMAX models were identified by the FROLS algorithm, together with a neural NARX model, and their one-step-ahead and free-run simulation residuals fed a filtered T^2 index, whose threshold was set at the 99% quantile of the fault-free validation simulations. Under normal operation, no degree-2 polynomial term reached the required error reduction ratio, so the NARX structure reduced to the ARX one, and the neural network did not improve the prediction, which indicates practically linear relations around the operating point. Even so, the residual-based detector achieved an average detection rate of 90.6% over the 17 detectable faults, against 86.7% for principal component analysis (PCA), with a false alarm rate of 1.3%, and outperformed PCA on the faults that change the dynamics without shifting the mean (10, 16, 19 and 20); the fusion of both detectors reached 93.3%, and the contribution of each subsystem to the alarm correctly pointed to the physical origin of faults such as 4, 11 and 14.**

***Keywords*—Fault detection, system identification, NARMAX, FROLS, Tennessee Eastman, residual analysis, data life cycle, MongoDB.**

Trabalho final da disciplina PCS5787 Ciência dos Dados e Big Data, Programa de Pós-Graduação em Engenharia Elétrica, Escola Politécnica da Universidade de São Paulo, 3º quadrimestre de 2026.

H. G. Santana (Nº USP 17981494, heitorgsantana@usp.br), H. B. Ferreira (Nº USP 12552502, helenabigliani@usp.br) e G. R. de Souza (Nº USP 18260136, gabrielrsouza04@hotmail.com) estão com a Escola Politécnica da Universidade de São Paulo, São Paulo, SP, Brasil.

I. Introdução

A detecção de falhas em processos industriais tornou-se um dos campos de aplicação mais estudados da ciência dos dados, pois as plantas modernas registram continuamente centenas de variáveis, e a quantidade de dados disponível supera em muito a capacidade de modelagem fenomenológica das equipes de engenharia [1], [2]. Nesse contexto, o processo Tennessee Eastman (TEP), proposto por Downs e Vogel [3] como problema de referência de controle de uma planta química completa, tornou-se o banco de ensaio padrão para comparar métodos de monitoramento, e o conjunto ampliado publicado por Rieth et al. [4], com milhões de registros de operação normal e sob 20 tipos de falha, permite avaliações estatisticamente mais robustas do que o conjunto original, que tinha uma única simulação por condição.

Os métodos mais difundidos para o TEP são multivariados e estáticos, como a análise de componentes principais (PCA) com os índices T^2 e SPE [2], que não exploram explicitamente a dinâmica entre as entradas e as saídas de cada parte do processo. Uma alternativa é a detecção por resíduos de modelos: identifica-se, com dados de operação normal, um modelo que prediz a saída de um subsistema a partir dos valores passados de suas entradas e saídas, e qualquer alteração dessa relação, provocada por uma falha, aparece como aumento do erro de predição. Os modelos NARMAX (*nonlinear autoregressive moving average with exogenous inputs*) e seus casos particulares, NARX e ARX, identificados

pelo algoritmo FROLS (*forward regression orthogonal least squares*), são adequados a esse fim porque produzem estruturas parcimoniosas e interpretáveis, selecionadas a partir dos dados pela taxa de redução de erro (ERR) [5], [6].

Este artigo descreve o projeto final do grupo na disciplina PCS5787, que aplicou essa abordagem aos dados de Rieth et al. [4] percorrendo todas as etapas do ciclo de vida dos dados [7], a saber: planejamento por um plano de gestão de dados, obtenção e armazenamento num banco NoSQL orientado a documentos, verificação de qualidade e análise exploratória, identificação dos modelos, detecção e isolamento de falhas, e publicação dos dados derivados e do código. As questões analíticas investigadas foram cinco, apresentadas na Seção III com as hipóteses correspondentes, e as principais contribuições são: (i) a evidência, sustentada por seleção de estrutura, por teste estatístico e pela comparação com uma rede neural, de que as relações entre as variáveis do TEP são praticamente lineares em operação normal; (ii) a comparação, com o mesmo nível de falsos alarmes, entre detectores por resíduos e a PCA, com intervalos de confiança obtidos por reamostragem; e (iii) um procedimento reprodutível, com código aberto e dados derivados publicados.

O trabalho foi dividido entre os integrantes da seguinte forma: Heitor Gonçalves Santana coordenou o projeto e foi responsável pela identificação dos modelos NARX/NARMAX e pelo detector por resíduos; Helena Bigliani Ferreira foi responsável pelo plano de gestão de dados, pela modelagem e carga do banco orientado a documentos e pela documentação e publicação no Zenodo; e Gabriel Rios de Souza foi responsável pela análise exploratória, pela linha de base PCA e pela avaliação estatística das métricas de detecção, tendo a redação do artigo sido feita em conjunto. O restante do texto está organizado assim: a Seção II descreve os dados e seu ciclo de vida; a Seção III, as questões e hipóteses; a Seção IV, os métodos; a Seção V, os resultados; a Seção VI, a publicação; e a Seção VII, as conclusões.

II. Dados e ciclo de vida

A. Conjunto de dados

O conjunto *Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation* [4] contém simulações em malha fechada do TEP com 52 variáveis de processo, das quais 41 são medidas (`xmeas_1` a `xmeas_41`) e 11 são manipuladas (`xmv_1` a `xmv_11`), amostradas a cada 3 min, além de três atributos de identificação, a saber: o tipo de falha (`faultNumber`, 0 para operação normal), o número da simulação (`simulationRun`) e o índice da amostra (`sample`). Foram usados os dois arquivos de treinamento, cada simulação tendo 500 amostras (25 h): o de operação normal, com 500 simulações (250 000 registros), e o de falhas, com 500 simulações para cada uma das 20 falhas (5 000 000 registros), nas quais o distúrbio é introduzido após a amostra 20, isto é, após 1 h de operação. Os arquivos, no formato RData, foram lidos por um leitor sequencial escrito pelo grupo, que converte as colunas para precisão simples à medida que são lidas, o que permitiu processar o arquivo de falhas, de cerca de 2,2 GB em memória em dupla precisão, num computador pessoal.

Tabela I
Etapas do ciclo de vida dos dados e ações do projeto

Etapa	Ação e artefato
Planejar	Plano de gestão de dados (fontes, formatos, licenças CC BY 4.0 e MIT, responsáveis).
Coletar	*Download* dos RData do Harvard Dataverse (DOI 10.7910/DVN/6C3JR1) e extração por leitor próprio.
Assegurar	Verificação de completude (500 amostras por simulação, índices contíguos), tipos e taxas de atualização.
Descrever	Dicionário de dados com descrição, unidade e período de cada variável, gravado no banco.
Preservar	Banco de documentos e extratos reconstruíveis pelo código; resultados publicados com versão (v1.0.0); dados brutos referenciados pelo DOI.
Descobrir e integrar	Consulta por condição e simulação; integração dos arquivos normal e com falhas num só esquema.
Analisar	Análise exploratória, identificação, detecção e isolamento (Seções IV e V).
Publicar	Depósito no Zenodo, comunidade do grupo de Big Data Analytics da Poli-USP (DOI 10.5281/zenodo.24532683).

B. Ciclo de vida dos dados

A Tabela I relaciona as etapas do ciclo de vida do DataONE [7], discutido na disciplina, às ações e aos artefatos do projeto. As decisões de planejamento, a saber: as fontes, os formatos, as licenças e o destino dos dados derivados, foram tomadas antes da análise e consolidadas no plano de gestão de dados publicado com o depósito, e a avaliação de qualidade e a descrição dos dados foram feitas na análise exploratória do grupo, cujos resultados orientaram a seleção dos atributos (Seção II-D).

C. Modelagem e armazenamento NoSQL

Os dados foram modelados num banco orientado a documentos, com quatro coleções: `variaveis`, que guarda o dicionário de dados; `simulacoes`, com um documento por simulação; `modelos`, com a estrutura, os parâmetros e a ERR de cada modelo identificado; e `metricas`, com os resultados de detecção por falha e detector. A unidade natural de consulta e de análise é a simulação, pois os modelos dinâmicos exigem séries contíguas e a separação entre treino, validação e teste é feita por simulação; por isso, cada documento de `simulacoes` reúne os metadados (`faultNumber`, `simulationRun`, conjunto, origem, DOI da fonte, amostra de início da falha) e a matriz 500×52 das variáveis, armazenada como bloco binário em precisão simples, o que evita 26 000 pares chave-valor por documento e mantém cada documento em cerca de 100 kB, bem abaixo do limite de 16 MB do MongoDB. Criou-se um índice sobre `faultNumber`, e as consultas da análise selecionam as simulações por condição e por faixa de `simulationRun`.

O banco foi implementado com a biblioteca Mongita, que reproduz a API do cliente PyMongo e grava os documentos em disco, de modo que o mesmo código opera sobre um servidor MongoDB trocando-se apenas a linha de conexão; essa escolha dispensou a instalação de um servidor durante o desenvolvimento, sem prejuízo da modelagem. Foram carregadas as 500 simulações normais e as simulações 426 a 500 de cada falha, que formam o conjunto de teste, totalizando 2 000 documentos e cerca de 210 MB.

D. Seleção e preparação dos atributos

Seguindo a análise exploratória do grupo, adotaram-se as seguintes decisões: (i) as 19 composições (`xmeas_23` a `xmeas_41`) foram excluídas, pois são atualizadas a cada 6 ou 15 min e permanecem constantes entre atualizações, o que introduz degraus artificiais incompatíveis com modelos a cada 3 min, restando 33 variáveis contínuas; (ii) todas as variáveis foram padronizadas pelo escore z, com média e desvio padrão calculados somente nas simulações normais de treino; e (iii) as simulações foram separadas em treino (normais 1 a 350), validação (normais 351 a 425) e teste (normais 426 a 500 e, para cada falha, as simulações 426 a 500), separação feita por `simulationRun`, e nunca por amostra, para evitar vazamento de informação entre amostras correlacionadas no tempo.

III. Questões analíticas e hipóteses

Formularam-se cinco questões, cada uma associada a uma hipótese verificável com os dados, a saber:

- **Q1**: há não linearidade relevante nas relações entre entradas e saídas dos subsistemas em operação normal? **H1**: termos polinomiais de grau 2 reduzem significativamente o erro de predição em relação ao modelo ARX.
- **Q2**: um modelo de ruído melhora a predição? **H2**: os termos de média móvel do NARMAX tornam os resíduos mais brancos e reduzem o erro quadrático médio (EQM) de predição.
- **Q3**: os resíduos de modelos identificados em operação normal detectam as falhas melhor que a PCA, com a mesma taxa de falsos alarmes? **H3**: o detector por resíduos supera a PCA nas falhas que alteram a dinâmica sem deslocar a média, como as aleatórias e as de agarramento de válvula.
- **Q4**: os resíduos indicam onde a falha ocorreu? **H4**: o subsistema com maior proporção de alarmes coincide com a origem física da falha.
- **Q5**: as falhas 3, 9 e 15, sem assinatura estatística na análise exploratória, são detectáveis? **H5**: nenhum dos detectores supera, nessas falhas, a taxa de falsos alarmes.

IV. Métodos

A. Subsistemas

Foram definidos cinco subsistemas de uma saída e múltiplas entradas (Tabela II), escolhidos segundo a análise de correlação cruzada da etapa exploratória, que indicou relações dinâmicas entre a válvula de vapor e a temperatura do *stripper*, memória longa na pressão do reator e uma relação entre a vazão de água de refrigeração e a temperatura do reator que muda quando o controle compensa um distúrbio. Evitaram-se os pares de nível e válvula, cuja correlação unitária e instantânea indica relação estática sem informação dinâmica. A Fig. 1 situa os cinco subsistemas no fluxograma do processo, que reúne o reator, o condensador, o separador vapor-líquido, o compressor de reciclo e o *stripper*.

Tabela II
Subsistemas identificados

	Saída y	Entradas u
S1	`xmeas_21`, temp. água do reator	`xmv_10` (vazão de água), `xmeas_9` (temp. do reator)
S2	`xmeas_18`, temp. do *stripper*	`xmv_9` (válvula de vapor), `xmeas_14` (fundo do separador)
S3	`xmeas_7`, pressão do reator	`xmv_6` (purga), `xmv_5` (reciclo), `xmeas_6` (alimentação)
S4	`xmeas_11`, temp. do separador	`xmv_11` (água do condensador), `xmeas_22` (temp. dessa água)
S5	`xmeas_20`, potência do compressor	`xmv_5`, `xmeas_5` (reciclo), `xmeas_7`

B. Modelos NARX e NARMAX

O modelo NARMAX polinomial descreve a saída como

$$y(k) = F^{\ell}\big[y(k-1),\dots,y(k-n_y),\,u(k),\dots,u(k-n_u),\\ e(k-1),\dots,e(k-n_e)\big] + e(k), \quad (1)$$

em que F^{ℓ} é um polinômio de grau ℓ e $e(k)$ é o resíduo; o NARX é o caso sem os termos de ruído, e o ARX, o caso linear ($\ell = 1$) [5]. Usaram-se $n_y = 3$, $n_u = 2$ e $n_e = 2$, incluindo-se o regressor $u(k)$, pois a detecção opera sobre amostras já medidas e várias relações do TEP são praticamente instantâneas na escala de 3 min, e adotou-se $\ell = 2$ para o NARX e o NARMAX. O polinômio F^{ℓ} não é escolhido de antemão: ele é a família de todos os monômios de grau até ℓ formados a partir do vetor de regressores

$$\boldsymbol{\varphi}(k) = \big[y(k-1)\;\; y(k-2)\;\; y(k-3)\;\; u_1(k)\;\; \cdots\;\; u_p(k-2)\big]^{\mathsf{T}}, \quad (2)$$

que tem $n = 9$ elementos nos subsistemas de duas entradas (S1, S2 e S4) e $n = 12$ nos de três (S3 e S5). Os polinômios candidatos de grau 1 e de grau 2 são, respectivamente,

$$F^1 = \theta_0 + \sum_{i=1}^{n} \theta_i\, \varphi_i(k), \quad (3)$$

$$F^2 = \theta_0 + \sum_{i=1}^{n} \theta_i\, \varphi_i(k) + \sum_{i=1}^{n}\sum_{j=i}^{n} \theta_{ij}\, \varphi_i(k)\, \varphi_j(k), \quad (4)$$

o que gera $1+n$ candidatos no ARX (10 ou 13) e $1+n+n(n+1)/2$ no NARX (55 ou 91). No NARMAX, acrescentam-se sete candidatos de ruído, a saber: $e(k-1)$, $e(k-2)$, $e(k-1)^2$, $e(k-1)e(k-2)$, $e(k-2)^2$, $y(k-1)e(k-1)$ e $y(k-1)e(k-2)$.

Escrevendo-se (1) na forma linear nos parâmetros, $\mathbf{y} = \mathbf{P}\boldsymbol{\theta} + \mathbf{e}$, em que as colunas $\mathbf{p}_i$ de $\mathbf{P}$ são os termos candidatos avaliados em todas as amostras de treino, o algoritmo FROLS [5], [6] ortogonaliza progressivamente os candidatos, $\mathbf{w}_i = \mathbf{p}_i - \sum_r \frac{\mathbf{q}_r^{\mathsf{T}}\mathbf{p}_i}{\mathbf{q}_r^{\mathsf{T}}\mathbf{q}_r}\mathbf{q}_r$, em relação aos termos $\mathbf{q}_r$ já escolhidos, e seleciona, a cada passo, o termo de maior taxa de redução de erro,

$$\mathrm{ERR}_i = \frac{(\mathbf{y}^{\mathsf{T}}\mathbf{w}_i)^2}{(\mathbf{y}^{\mathsf{T}}\mathbf{y})(\mathbf{w}_i^{\mathsf{T}}\mathbf{w}_i)}, \quad (5)$$

que mede a fração da energia da saída explicada pelo termo. A seleção termina quando a ERR do melhor candidato restante é menor que $\rho = 10^{-4}$, com limite de 20 termos, e os parâmetros são então estimados por mínimos quadrados. As

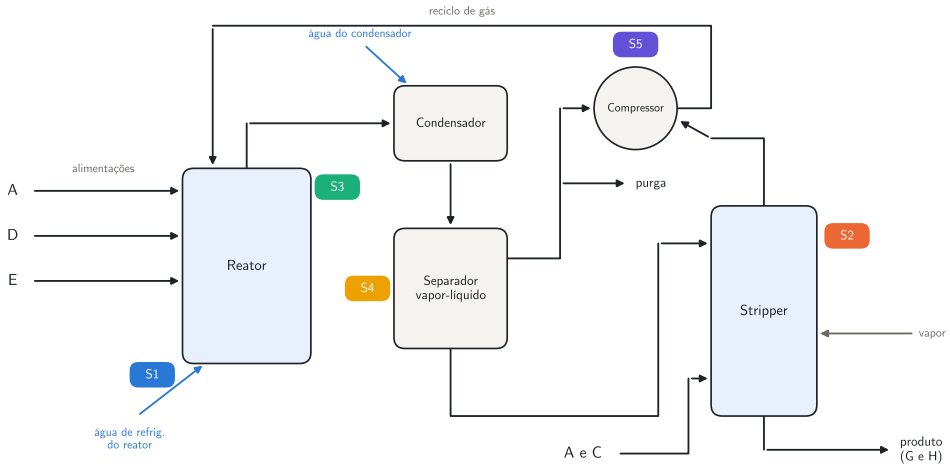


Fig. 1. Esquema simplificado do TEP, redesenhado a partir de Downs e Vogel [3], com a localização dos subsistemas S1 a S5: S1, refrigeração do reator; S2, *stripper*; S3, pressão do reator; S4, separador e condensador; S5, compressor.

Tabela III
SELEÇÃO DE ESTRUTURA PELO FROLS NO SUBSISTEMA S2 ($\rho = 10^{-4}$)

Passo	Termo	ERR (%)	Acumulada (%)
1	$y(k-1)$	98,66	98,66
2	$y(k-2)$	0,317	98,98
3	$u_1(k)$	0,176	99,15
4	$y(k-3)$	0,053	99,21
5	$u_1(k-2)$	0,039	99,25
–	$y(k-1)\,u_1(k-1)$ *	0,0005	–

* Melhor candidato quadrático, não selecionado.

regressões foram montadas dentro de cada simulação e depois empilhadas, o que evita termos que misturem o fim de uma simulação com o início da seguinte, e a implementação própria foi verificada contra a biblioteca SysIdentPy [8], que, numa configuração comum às duas, selecionou os mesmos termos na mesma ordem, com ERR e parâmetros idênticos até a sexta casa decimal. A Tabela III ilustra a seleção em S2: cada candidato foi avaliado nas 173 950 amostras de treino, o termo $y(k-1)$ explica sozinho 98,66% da energia da saída, seguem-se $y(k-2)$, $u_1(k)$, $y(k-3)$ e $u_1(k-2)$, e o processo termina com 99,25%, pois nenhum candidato restante supera o limiar; o melhor termo quadrático, $y(k-1)\,u_1(k-1)$, teria ERR de 0,0005%, vinte vezes abaixo de ρ. A constante θ_0 não foi selecionada em nenhum subsistema, como se espera com dados padronizados.

No NARMAX, os termos de processo selecionados no NARX são fixados, e o FROLS escolhe, com $\rho = 10^{-4}$ e no máximo quatro termos, os de ruído entre $e(k-1)$, $e(k-2)$, seus produtos e seus produtos com $y(k-1)$; os parâmetros são estimados por mínimos quadrados estendidos, alternando-se, em dez iterações, a estimação e o recálculo recursivo dos resíduos [5]. Como validação do modelo de ruído, ele só é mantido se reduzir o EQM de predição nas simulações de validação; caso contrário, o NARMAX coincide com o NARX. Para avaliar não linearidades não polinomiais, identificou-se também um NARX neural, perceptron multicamadas (MLP) com uma camada oculta de 32 neurônios de ativação tangente hiperbólica e saída linear, que recebe os mesmos 9 ou 12 regressores de (2) e tem, portanto, de 353 a 449 parâmetros [9]. O treino usou o otimizador Adam, com lotes de 512 amostras e regularização L2 de 10^{-4}, e parada antecipada: 10% das amostras de treino foram reservadas para acompanhar o erro, e o treino era interrompido após seis épocas sem melhora, com limite de 60 épocas, o que ocorreu entre 19 e 31 épocas, conforme o subsistema.

C. Validação dos modelos

Além do coeficiente de determinação, calculado nas simulações de validação e de teste para verificar a generalização, os modelos foram submetidos aos testes de correlação do resíduo propostos por Billings [5], que verificam se resta no resíduo informação que o modelo deveria ter capturado, a saber: a autocorrelação do resíduo de predição um passo à frente,

$$\rho_e(\tau) = \frac{\sum_k [e(k)-\bar{e}]\,[e(k-\tau)-\bar{e}]}{\sum_k [e(k)-\bar{e}]^2}, \qquad \tau = 1,\ldots,10, \tag{6}$$

que deve ser nula para $\tau \geq 1$ se o resíduo for branco, isto é, se o modelo tiver capturado toda a dinâmica previsível; e a correlação cruzada normalizada $\rho_{ue}(\tau)$ entre o resíduo $e(k)$ e cada entrada $u_j(k-\tau)$, com $\tau = 0,\ldots,10$, que deve ser nula se nenhuma parte da saída ainda puder ser explicada pelas entradas. Os coeficientes foram calculados em cada uma das 75 simulações normais de teste e depois promediados, e a média foi comparada com a banda de 95% de uma única simulação, $\pm 1{,}96/\sqrt{N} = \pm 0{,}088$ para $N = 497$ amostras, critério conservador, pois a banda da média seria mais estreita. Os testes de correlação não lineares do mesmo autor não foram aplicados, dada a evidência de linearidade discutida na Seção V.

D. Resíduos e detector

Para cada modelo calcularam-se dois resíduos: o de predição um passo à frente (OSA), $e(k) = y(k) - \hat{y}(k\,|\,k-1)$, em

que as saídas passadas são as medidas, e o de simulação livre, em que as saídas passadas são as próprias predições do modelo a partir das três primeiras amostras de cada simulação. O primeiro responde rapidamente a mudanças, mas tende a absorver desvios persistentes, pois o modelo é realimentado pela saída medida; o segundo acumula o efeito de desvios persistentes da relação entre entradas e saída.

Os resíduos dos cinco subsistemas formam o vetor $\mathbf{r}(k)$, e o índice de detecção é a estatística

$$D(k) = \frac{1}{W} \sum_{j=0}^{W-1} \big[\mathbf{r}(k-j) - \boldsymbol{\mu}\big]^{\mathsf{T}} \mathbf{S}^{-1} \big[\mathbf{r}(k-j) - \boldsymbol{\mu}\big], \quad (7)$$

em que $\boldsymbol{\mu}$ é o vetor das médias de cada resíduo e $\mathbf{S}$ a matriz de covariância amostral dos resíduos, de dimensão 10×10 no vetor completo, ambos estimados nas simulações normais de validação (351 a 425), que não participam nem da identificação nem do teste. A subtração de $\boldsymbol{\mu}$ centra os resíduos, e a inversa de $\mathbf{S}$, calculada como pseudoinversa, normaliza as escalas e desconta a correlação entre os resíduos, de modo que resíduos de variâncias diferentes pesem igualmente e resíduos correlacionados, como o OSA e o de simulação livre de um mesmo subsistema, não sejam contados duas vezes; $D(k)$ é, assim, a média móvel do quadrado da distância de Mahalanobis. A janela de $W = 10$ amostras (30 min) é a janela da média móvel, que, como mostrou a análise exploratória, torna detectáveis as falhas que alteram apenas a dispersão. O alarme é declarado quando $D(k)$ excede o limiar τ, quantil de 99% de D nas simulações de validação, o que corresponde a uma taxa nominal de falsos alarmes de 1%. Consideraram-se os vetores de resíduos OSA, de simulação livre e a concatenação dos dois (dez resíduos).

O mesmo índice detecta tanto falhas que deslocam a média quanto falhas que alteram a dispersão. Se, durante uma falha, os resíduos passam a ter média $\boldsymbol{\mu}+\boldsymbol{\delta}$ e covariância $\mathbf{S}_f$, o valor esperado de cada parcela de (7) é

$$\mathrm{E}_f\big[d^2(k)\big] = \mathrm{tr}\big(\mathbf{S}^{-1}\mathbf{S}_f\big) + \boldsymbol{\delta}^{\mathsf{T}}\mathbf{S}^{-1}\boldsymbol{\delta}, \quad (8)$$

em que $d^2(k)$ é o quadrado da distância de Mahalanobis no instante k: o primeiro termo mede a alteração da dispersão e o segundo, o deslocamento da média, e, em operação normal, $\boldsymbol{\delta} = \mathbf{0}$ e $\mathbf{S}_f = \mathbf{S}$, de modo que o valor esperado se reduz ao número m de resíduos. A média móvel não altera esse valor esperado, mas reduz a variância do índice, o que separa as distribuições normal e em falha quando o efeito é pequeno e persistente, ao custo de um atraso de até W amostras em falhas abruptas.

E. Linha de base PCA e fusão

A PCA foi ajustada às 33 variáveis padronizadas das simulações normais de treino, retendo-se 17 componentes, que explicam 92,0% da variância, número indicado pela análise exploratória; calcularam-se os índices T^2 e SPE [2], filtrados com a mesma janela W, e combinou-se os dois num único índice, o maior entre T^2/τ_{T^2} e $\mathrm{SPE}/\tau_{\mathrm{SPE}}$, recalibrado para 1% de falsos alarmes na validação. A fusão entre o detector por resíduos e a PCA seguiu a mesma regra, de modo que todos os detectores comparados têm o mesmo limiar nominal.

F. Métricas

Para cada falha e detector, calcularam-se, nas 75 simulações de teste: a taxa de detecção (FDR), fração das amostras posteriores à introdução da falha (amostras 21 a 500) em alarme; a taxa de falsos alarmes (FAR), fração de amostras em alarme nas 75 simulações normais de teste; e o atraso de detecção, tempo entre a introdução da falha e o início da primeira sequência de três alarmes consecutivos, resumido pela mediana. Os intervalos de confiança de 95% da FDR média foram obtidos por reamostragem *bootstrap* das simulações, com 1 000 réplicas [10], e as diferenças de EQM entre modelos foram avaliadas pelo teste pareado de Wilcoxon [11] sobre o EQM de cada simulação normal de teste. A FDR média é apresentada sobre as 17 falhas detectáveis, excluídas as falhas 3, 9 e 15 (Q5).

G. Estratégia de visualização

As figuras do artigo e da apresentação foram projetadas segundo o quadro de Munzner [13], que decompõe cada visualização em três perguntas, a saber: o que se mostra, isto é, a abstração dos dados (tabelas com atributos-chave e atributos-valor, de tipo categórico, ordenado ou quantitativo); por que se mostra, isto é, a tarefa abstrata do leitor (descobrir, apresentar, comparar, identificar, em relação a alvos como tendências, distribuições e discrepâncias); e como se mostra, isto é, o idioma visual e os canais que codificam cada atributo. Adotaram-se, em particular, os dois princípios que Munzner enuncia para a escolha dos canais: o da expressividade, segundo o qual atributos categóricos são codificados por canais de identidade, como o matiz e a forma, e atributos ordenados por canais de magnitude, como a posição, o comprimento e a luminância; e o da eficácia, segundo o qual o atributo mais importante de cada figura recebe o canal mais preciso, a posição numa escala comum alinhada. Seguiram-se ainda três estratégias do mesmo autor para lidar com a complexidade, a saber: facetar em vistas justapostas (*small multiples*) em vez de sobrepor muitas séries, reordenar as linhas e colunas de matrizes para revelar grupos e reduzir os dados por agregação, sempre acompanhada de uma medida de incerteza. A Tabela IV resume a análise de cada figura.

As cores seguem uma única convenção em todas as figuras e nos slides: azul para os resíduos, laranja para a PCA e verde para a fusão, nos gráficos de detecção, e laranja para termos autorregressivos e azul para termos de entrada, nos de modelos; a paleta foi verificada quanto à separação para leitores com deficiência de visão de cores, e a identidade do detector é codificada também pela forma do marcador, de modo que não depende apenas da cor. Evitaram-se eixos duplos, gráficos tridimensionais e barras em escala logarítmica, e os números seguem a vírgula decimal.

V. Resultados e discussão

A. Identificação (Q1 e Q2)

A Tabela V resume os modelos identificados, e a Fig. 3 detalha a estrutura e os parâmetros de cada um. Em nenhum subsistema o FROLS selecionou termos de grau 2 com

Tabela IV
ANÁLISE DAS FIGURAS SEGUNDO O QUADRO O QUÊ/POR QUÊ/COMO DE MUNZNER [13]

Figura	O quê (abstração dos dados)	Por quê (tarefa)	Como (idioma e canais)
Fig. 3	Por subsistema, lista de termos (categórico, ordenado pela seleção) com coeficiente (quantitativo com sinal) e ERR	Apresentar a estrutura; comparar importância e sinal dos termos	Barras divergentes a partir de zero (posição e comprimento), matiz para o tipo de termo, ERR anotada; cinco vistas justapostas com o mesmo eixo
Fig. 2	Um valor quantitativo por subsistema	Comparar com o limiar de seleção	Pontos em escala logarítmica, sem barras (cuja base seria arbitrária), linha de referência
Fig. 4	Detector (categórico) com FDR média e FAR (quantitativos)	Ordenar e comparar, com incerteza	Gráfico de pontos ordenado pelo valor, barras de IC 95%, dois painéis alinhados pela mesma chave
Fig. 5	Tabela com duas chaves (falha, detector) e um valor (FDR)	Comparar PCA e resíduos em cada falha; identificar onde há ganho	Gráfico de halteres: posição vertical em escala comum, segmento que codifica a diferença, matiz e forma para o detector; facetas pelas classes de falha da EDA
Fig. 7	Séries temporais (quantitativo por tempo ordenado)	Descobrir o comportamento temporal; comparar com o limiar	Linhas em escala logarítmica comum, linha de referência, vistas justapostas
Fig. 6	Matriz com duas chaves (falha, subsistema) e um valor (fração de alarmes)	Identificar a origem; descobrir grupos de falhas	Mapa de calor com luminância sequencial de um só matiz e colunas reordenadas por assinatura

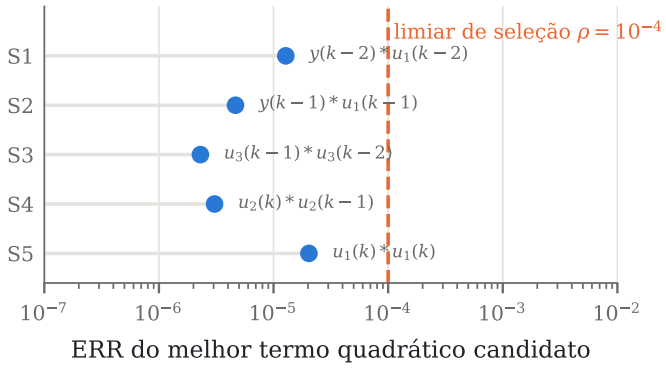


Fig. 2. ERR do melhor termo quadrático candidato em cada subsistema, calculada após a ortogonalização em relação aos termos lineares selecionados, e limiar de seleção do FROLS (linha tracejada); ao lado de cada ponto, o termo correspondente.

$\rho = 10^{-4}$: a estrutura NARX resultou idêntica à ARX, termo a termo, e só com $\rho = 10^{-5}$ apareceram termos quadráticos, em S1 e S5, sem alteração do coeficiente de determinação R^2 na terceira casa decimal. A rede neural, que pode representar não linearidades de outra natureza, obteve R^2 igual ao do ARX na predição OSA, com diferenças apenas na quarta casa decimal, e igual ou inferior na simulação livre, chegando a 0,563 contra 0,613 em S5. Essa coincidência não decorre de erro, mas do que a rede aprendeu: um ajuste linear às suas próprias predições, sobre os mesmos regressores, explica de 99,6% a 99,99% da variância da sua saída, e mais de 92% das ativações dos neurônios ocultos ficam na faixa $|z| < 1$, em que a tangente hiperbólica é quase linear; a rede convergiu, portanto, para praticamente o mesmo mapeamento linear do ARX, o que constitui uma evidência de linearidade independente do FROLS. A Fig. 2 mostra a evidência de forma direta: depois de incluídos os termos lineares, o melhor candidato quadrático de cada subsistema explica no máximo cerca de 2×10^{-5} da energia da saída, valor de cinco a quarenta vezes menor que o limiar, conforme o subsistema. Rejeita-se, portanto, a hipótese H1: em torno do ponto de operação, com as pequenas variações provocadas pelos distúrbios aleatórios da operação normal, as relações entre as variáveis do TEP são praticamente lineares, e o princípio da parcimônia conduz ao modelo ARX.

As equações completas dos modelos estão na Tabela VI. Os modelos têm boa capacidade preditiva em S2, S3 e S5 (R^2 OSA acima de 0,95) e menor em S1 e S4, cujas saídas têm parcela importante de ruído de medição; em S1, por exemplo, o modelo selecionado, em que u_1 é a vazão de água de refrigeração e u_2 a temperatura do reator, tem como termos mais relevantes justamente as entradas no instante atual e no anterior, coerentes com a física do trocador de calor. Nos demais subsistemas, a Fig. 3 mostra o padrão oposto: o termo $y(k-1)$ responde sozinho por 59 a 99% da ERR, o que indica dinâmicas lentas em relação ao período de amostragem, e os termos de entrada, embora de ERR pequena, são os que carregam a relação de causa e efeito cuja alteração o detector procura. O modelo de ruído foi mantido em S1 e S2 e rejeitado pela validação em S3, S4 e S5; em S2, o termo $e(k-1)$ reduziu o EQM em 5,3% ($p < 10^{-13}$) e levou a autocorrelação do resíduo no atraso 1 de $-0{,}06$ para praticamente zero, ao passo que, em S1, a redução, embora significativa, foi de apenas 0,2%. Aceita-se H2 apenas parcialmente: o modelo de ruído branqueia o resíduo onde há correlação remanescente, mas a melhora é pequena e, em S3, S4 e S5, não se confirma na validação. A seleção de estrutura também descartou entradas: a vazão de fundo do separador (`xmeas_14`) não entrou no modelo de S2, nem a vazão de alimentação (`xmeas_6`) no de S3, por não reduzirem o erro; em S4, o termo $y(k-2)$ não foi selecionado, e a temperatura da água do condensador é a entrada dominante, com 20,6% da ERR; e S5, com dez termos, é o único que usa as três entradas, embora a pressão do reator entre com peso pequeno.

A Tabela VII resume a validação. O R^2 de validação e o de teste diferem no máximo 0,003, com modelos de 5 a 10 termos, o que afasta o sobreajuste. Nos testes de correlação, S1 e S3 têm resíduo branco e não correlacionado com as entradas, dentro da banda de 95%; S2 fica no limite da banda, e S4 e S5 a excedem na correlação entre resíduo e entradas (0,209 e 0,183), o que indica dinâmica de entrada não capturada com dois atrasos, e S4 apresenta ainda autocorrelação negativa no atraso 1 ($-0{,}153$), típica de ruído de medição não modelado. O R^2 de 0,52 de S1, por sua vez, não decorre de deficiência do modelo, pois o resíduo é branco e nenhuma das estruturas avaliadas, NARMAX ou rede neural, o reduz: ele reflete a

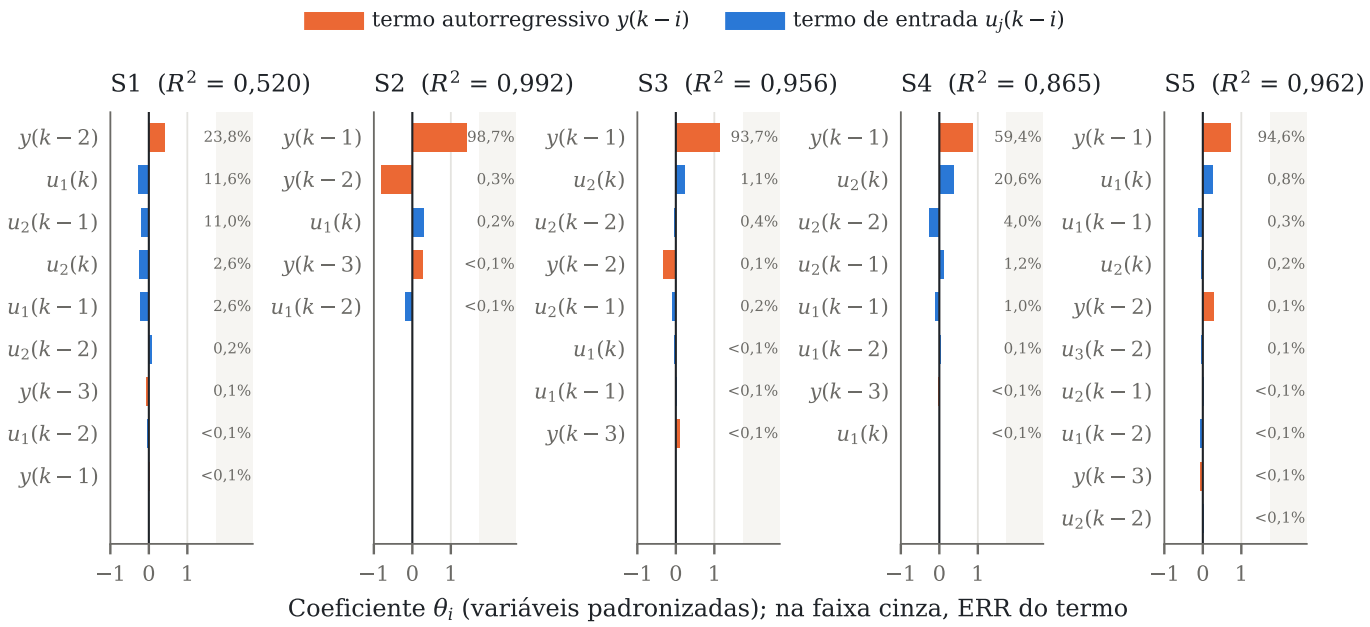


Fig. 3. Modelos ARX (iguais aos NARX polinomiais) dos cinco subsistemas: coeficientes de cada termo, em variáveis padronizadas, na ordem em que o FROLS os selecionou, e, na faixa cinza, a ERR de cada termo; no título, o R^2 de predição um passo à frente no teste normal. As entradas u_j de cada subsistema são as da Tabela II, na ordem em que lá aparecem.

Tabela V
MODELOS IDENTIFICADOS: NÚMERO DE TERMOS E SOMA DAS ERR DO ARX (IGUAL AO NARX), R^2 NO TESTE NORMAL EM PREDIÇÃO OSA E EM SIMULAÇÃO LIVRE, AUTOCORRELAÇÃO DO RESÍDUO OSA NO ATRASO 1 ($\rho_e(1)$), RESULTADO DO NARMAX (TERMOS; REDUÇÃO DO EQM OSA; TESTE DE WILCOXON) E R^2 DO NARX NEURAL (OSA/SIMULAÇÃO)

	ARX = NARX polinomial					NARMAX	NARX neural
Subsistema	Termos	$\sum$ERR	R^2 OSA	R^2 sim.	$\rho_e(1)$		R^2 OSA/sim.
S1	9	0,5191	0,520	0,419	-0,00	10; 0,2%; $p < 10^{-3}$	0,520/0,420
S2	5	0,9925	0,992	0,939	-0,06	6; 5,3%; $p < 10^{-13}$	0,992/0,935
S3	8	0,9560	0,956	0,686	-0,02	rejeitado	0,956/0,666
S4	8	0,8645	0,865	0,409	-0,15	rejeitado	0,864/0,407
S5	10	0,9622	0,962	0,613	-0,02	rejeitado	0,962/0,563

Tabela VI
EQUAÇÕES DOS MODELOS IDENTIFICADOS, EM VARIÁVEIS PADRONIZADAS; AS ENTRADAS u_j SÃO AS DA TABELA II, NA ORDEM EM QUE LÁ APARECEM, E S3, S4 E S5 TÊM APENAS O MODELO ARX, POIS O MODELO DE RUÍDO FOI REJEITADO NA VALIDAÇÃO

Modelo	Equação
S1, ARX	$y(k) = 0{,}016\,y(k-1) + 0{,}411\,y(k-2) - 0{,}054\,y(k-3) - 0{,}274\,u_1(k) - 0{,}218\,u_1(k-1) - 0{,}041\,u_1(k-2) - 0{,}246\,u_2(k) - 0{,}204\,u_2(k-1) + 0{,}072\,u_2(k-2)$
S2, ARX	$y(k) = 1{,}420\,y(k-1) - 0{,}808\,y(k-2) + 0{,}280\,y(k-3) + 0{,}291\,u_1(k) - 0{,}185\,u_1(k-2)$
S2, NARMAX	$y(k) = 1{,}906\,y(k-1) - 1{,}444\,y(k-2) + 0{,}505\,y(k-3) + 0{,}234\,u_1(k) - 0{,}200\,u_1(k-2) - 0{,}578\,e(k-1)$
S3, ARX	$y(k) = 1{,}143\,y(k-1) - 0{,}319\,y(k-2) + 0{,}088\,y(k-3) - 0{,}040\,u_1(k) + 0{,}028\,u_1(k-1) + 0{,}229\,u_2(k) - 0{,}104\,u_2(k-1) - 0{,}046\,u_2(k-2)$
S4, ARX	$y(k) = 0{,}866\,y(k-1) - 0{,}027\,y(k-3) + 0{,}019\,u_1(k) - 0{,}100\,u_1(k-1) + 0{,}039\,u_1(k-2) + 0{,}387\,u_2(k) + 0{,}118\,u_2(k-1) - 0{,}262\,u_2(k-2)$
S5, ARX	$y(k) = 0{,}720\,y(k-1) + 0{,}284\,y(k-2) - 0{,}071\,y(k-3) + 0{,}259\,u_1(k) - 0{,}125\,u_1(k-1) - 0{,}061\,u_1(k-2) - 0{,}050\,u_2(k) + 0{,}028\,u_2(k-1) + 0{,}013\,u_2(k-2) - 0{,}027\,u_3(k-2)$

baixa razão sinal-ruído de um circuito fortemente regulado, em que a temperatura do reator, uma das entradas, varia com desvio padrão de 0,019 °C, da mesma ordem da resolução de 0,01 °C com que é registrada. Para a detecção, essas limitações são aceitáveis, porque o detector não exige resíduos brancos, mas um comportamento conhecido em operação normal, que $\boldsymbol{\mu}$, $\mathbf{S}$ e o limiar calibrado na validação incorporam, inclusive a correlação remanescente; coerentemente com ela, a taxa de falsos alarmes medida no teste, de 1,1% a 1,4%, ficou um pouco acima do valor nominal de 1%.

B. *Detecção (Q3 e Q5)*

A Tabela VIII compara os detectores. Com o mesmo limiar nominal, as taxas de falsos alarmes no teste ficaram entre

Tabela VII
VALIDAÇÃO DOS MODELOS ARX NAS 75 SIMULAÇÕES NORMAIS DE TESTE: NÚMERO DE TERMOS, R^2 OSA NA VALIDAÇÃO E NO TESTE, R^2 EM SIMULAÇÃO LIVRE, TESTES DE CORRELAÇÃO DO RESÍDUO (ATRASOS 1 A 10 PARA ρ_e E 0 A 10 PARA ρ_{ue}; BANDA DE 95%: $\pm$0,088) E AVALIAÇÃO

Sub.	Termos	R^2 val.	R^2 teste	R^2 sim.	$\rho_e(1)$	$\max \lvert\rho_e\rvert$	$\max \lvert\rho_{ue}\rvert$	Avaliação
S1	9	0,517	0,520	0,419	−0,001	0,060	0,069	Adequado; baixa razão sinal-ruído
S2	5	0,991	0,992	0,939	−0,056	0,103	0,103	Adequado; correlação marginal
S3	8	0,954	0,956	0,686	−0,021	0,062	0,057	Adequado
S4	8	0,862	0,865	0,409	−0,153	0,153	0,209	Aceitável; dinâmica residual
S5	10	0,961	0,962	0,613	−0,021	0,125	0,183	Aceitável; dinâmica residual

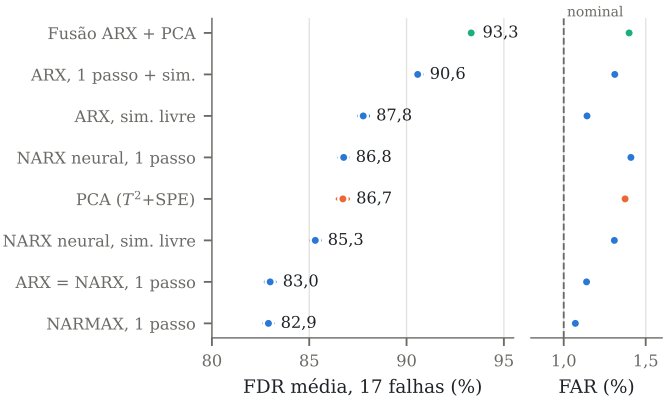


Fig. 4. FDR média nas 17 falhas detectáveis, com intervalo de confiança de 95% por *bootstrap* (menor que o marcador na maioria dos casos), e FAR no teste normal, para os oito detectores ordenados pela FDR; a linha tracejada marca a FAR nominal de 1%.

Tabela VIII
DESEMPENHO MÉDIO DOS DETECTORES: FDR MÉDIA NAS 17 FALHAS DETECTÁVEIS, FAR NO TESTE NORMAL, ATRASO MEDIANO E NÚMERO DE FALHAS COM FDR $\geq$ 50%, PARA AS JANELAS $W = 10$ E $W = 1$

	$W = 10$				$W = 1$	
Detector	FDR (%)	FAR (%)	Atr. (min)	N_{50}	FDR	FAR
PCA (T^2+SPE)	86,7	1,37	33	17	77,9	1,13
ARX = NARX, 1 passo	83,0	1,14	27	16	65,9	1,04
NARMAX, 1 passo	82,9	1,07	27	16	66,1	1,06
NARX neural, 1 passo	86,8	1,41	27	16	71,3	1,16
ARX, sim. livre	87,8	1,14	42	16	82,8	1,01
NARX neural, sim. livre	85,3	1,31	42	16	77,9	1,16
ARX, 1 passo + sim.	90,6	1,31	30	16	86,8	1,07
Fusão ARX + PCA	93,3	1,40	30	17	89,6	1,07

1,1% e 1,4%, próximas do 1% projetado, de modo que a comparação das taxas de detecção é equilibrada. Os detectores baseados apenas em resíduos OSA alcançaram FDR média pouco inferior à da PCA (83,0% contra 86,7%), mas com menor atraso mediano (27 contra 33 min); o resíduo de simulação livre superou a PCA (87,8%), e a concatenação dos dois tipos de resíduo alcançou 90,6%, combinando a rapidez de um com a sensibilidade a desvios persistentes do outro. A fusão com a PCA atingiu 93,3%, a maior taxa entre os detectores avaliados, com atraso mediano de 30 min. A Fig. 4 mostra que os intervalos de confiança das médias são estreitos, de no máximo $\pm$0,3 ponto percentual, de modo que as diferenças de ordenação entre os detectores são estatisticamente distinguíveis, exceto entre a PCA e o NARX neural OSA e entre o ARX e o NARMAX OSA, cujos intervalos se sobrepõem.

A Fig. 5 e a Tabela IX detalham o resultado por falha; a figura agrupa as falhas nas três classes da análise exploratória, o que torna visível que os maiores segmentos, isto é, as maiores diferenças entre PCA e resíduos, se concentram na classe das falhas que alteram a dispersão. As falhas que deslocam a média de alguma variável (1, 2, 4, 6, 8, 12, 14, 17 e 18) são detectadas pela PCA e pelos resíduos concatenados com FDR acima de 88%, com uma exceção instrutiva: na falha 1, o resíduo OSA isolado detecta apenas 62,1% das amostras, pois, após o transitório, o modelo realimentado pela saída medida absorve o desvio persistente, que o resíduo de simulação livre continua a acusar (98,1%). A diferença entre os métodos está nas demais falhas. Nas falhas 10, 16, 19 e 20, a primeira uma variação aleatória da temperatura da alimentação de C e as demais de tipo não divulgado por Downs e Vogel [3], a vantagem dos resíduos é nítida e os intervalos de confiança não se sobrepõem: na falha 19, a FDR passa de 57,6% (IC 95%: 55,6 a 59,6%) na PCA para 98,7% (98,6 a 98,9%) com os resíduos concatenados; na 16, de 62,5% para 93,0%; na 10, de 69,2% para 89,0%; e na 20, de 72,3% para 90,8%. Essas falhas alteram a variabilidade e a dinâmica do processo sem deslocar a média, e a média móvel do índice T^2 dos resíduos converte o aumento de variância das inovações em alarme, o que confirma H3. A decomposição (8), calculada com todas as amostras posteriores à falha nas 75 simulações de teste, torna essa distinção quantitativa: na falha 4, o termo de média vale 88,7 e o de dispersão, 9,8, igual ao normal; nas falhas 10, 11, 14, 16 e 19, o termo de média é inferior a 1, e a assinatura está toda no termo de dispersão, de 45 a 287; e, nas falhas 3, 9 e 15, os dois termos somam de 10,2 a 10,9, praticamente o valor normal $m = 10$. A Fig. 7 ilustra o comportamento na falha 19: o índice da PCA oscila em torno do limiar, ao passo que o dos resíduos OSA permanece de 1,5 a 7 vezes acima dele durante toda a falha. Na falha 4, ao contrário, a mudança de média é visível aos dois detectores, que cruzam o limiar praticamente na primeira amostra em falha, com atraso mediano de 3 min.

A PCA, por sua vez, é superior nas falhas 5 e 7, ambas em degrau, que o sistema de controle compensa: na falha 7, perda de pressão no cabeçalho da alimentação de C, o controle provavelmente restabelece as relações dos subsistemas modelados, que deixam de apresentar resíduo após o transitório, enquanto a mudança permanente nas vazões e aberturas de válvula continua visível para a PCA, que observa as 33 variáveis; a FDR dos resíduos OSA nessa falha é de

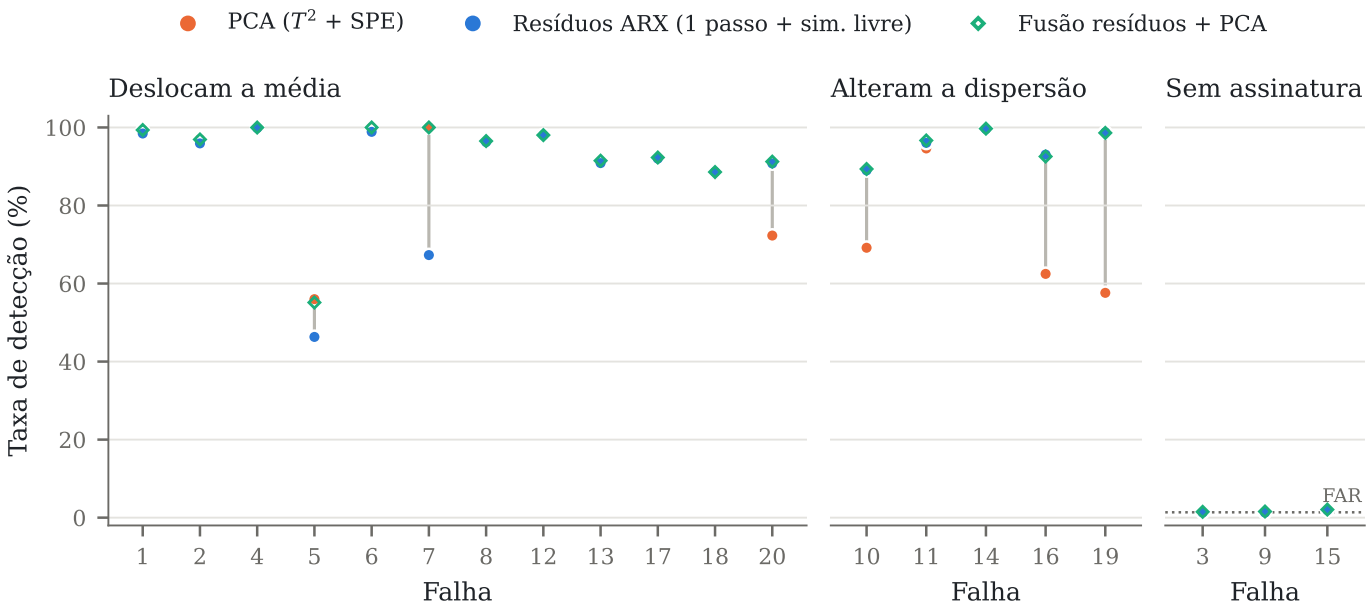


Fig. 5. Taxa de detecção por falha nas 75 simulações de teste ($W = 10$), agrupada pelas classes de falha da análise exploratória; o segmento cinza liga a PCA aos resíduos ARX, e seu comprimento é o ganho ou a perda de um em relação ao outro; a linha pontilhada, no último painel, marca a FAR.

Tabela IX
FDR (%) POR FALHA NAS SIMULAÇÕES DE TESTE ($W = 10$), INTERVALO DE CONFIANÇA DE 95% DA FUSÃO E ATRASO MEDIANO DE DETECÇÃO (MIN) DA PCA E DA FUSÃO; EM NEGRITO, A MAIOR FDR DE CADA FALHA DETECTÁVEL

Falha	PCA	ARX OSA	NARMAX OSA	ARX sim.	ARX OSA+sim.	Fusão	IC 95% fusão	Atraso PCA	Atraso fusão
1	**99,3**	62,1	55,6	98,1	98,5	**99,4**	99,3–99,4	12	12
2	**97,0**	95,1	95,1	96,0	95,9	**96,9**	96,8–97,0	48	48
3	1,5	1,2	1,1	1,1	1,3	1,5	0,9–2,2	–	–
4	**100,0**	**100,0**	**100,0**	99,8	99,9	**100,0**	100,0–100,0	3	3
5	**56,0**	31,2	31,0	46,7	46,3	55,1	53,1–57,3	9	9
6	**100,0**	99,5	99,5	98,7	98,9	**100,0**	100,0–100,0	3	3
7	**100,0**	54,6	54,4	66,9	67,3	**100,0**	100,0–100,0	3	3
8	**96,5**	94,9	94,9	96,1	96,3	**96,5**	96,3–96,8	51	54
9	1,6	1,3	1,2	1,2	1,3	1,6	1,0–2,3	–	–
10	69,2	68,5	68,9	88,4	89,0	**89,4**	88,1–90,5	114	69
11	94,6	95,7	96,0	94,1	96,1	**96,7**	96,2–97,1	30	27
12	97,9	98,1	**98,2**	98,0	98,0	98,0	97,8–98,2	30	30
13	91,4	89,2	88,7	90,9	90,9	**91,5**	90,8–92,2	120	120
14	99,7	**99,8**	**99,8**	99,5	99,6	99,7	99,7–99,7	6	6
15	1,9	2,5	2,5	1,6	1,9	2,1	1,3–3,0	–	–
16	62,5	78,7	83,1	91,9	**93,0**	92,6	91,7–93,3	63	42
17	**92,3**	**92,3**	**92,3**	92,0	92,0	**92,3**	91,5–93,0	114	117
18	88,0	88,3	88,3	**88,6**	**88,6**	**88,6**	87,4–89,7	171	159
19	57,6	**99,2**	**99,2**	55,8	98,7	98,6	98,5–98,7	33	18
20	72,3	63,7	64,4	90,6	90,8	**91,2**	90,5–92,0	123	123

54,6%, contra 100% da PCA. Esse resultado mostra que os dois detectores são complementares, o que explica o ganho da fusão, que obteve a maior FDR, ou empatou com a maior, em 12 das 17 falhas detectáveis e ficou a menos de 1 ponto percentual da melhor nas outras cinco.

Nas falhas 3, 9 e 15, todos os detectores tiveram FDR entre 1,1% e 2,5%, valores da mesma ordem da taxa de falsos alarmes, o que confirma H5 e o resultado da análise exploratória, segundo o qual essas falhas não alteram nem a média nem a dispersão das variáveis: o maior deslocamento de média foi de 0,26 desvio padrão e a maior razão entre desvios padrão, de 1,10. A Tabela X resume o efeito dessas falhas, e a explicação está na sua natureza. Nas falhas 3 e 9, degrau e variação aleatória da temperatura da alimentação D, a variável perturbada não é medida, a corrente D traz pouca entalpia em comparação com o calor de reação e com a inércia térmica do reator, e a malha de temperatura do reator compensa o restante com uma correção menor que suas oscilações normais. Na falha 15, agarramento da válvula de água de refrigeração do condensador, o atrito estático só se manifesta quando a válvula precisa se mover além da zona morta, o que, perto do ponto de operação, ocorre pouco; o único efeito observado foi um aumento de 9,5% no desvio padrão da temperatura da água de saída do condensador (xmeas_22), sem deslocamento de média e menor que o ruído do subsistema S4. Quando as distribuições em falha e em

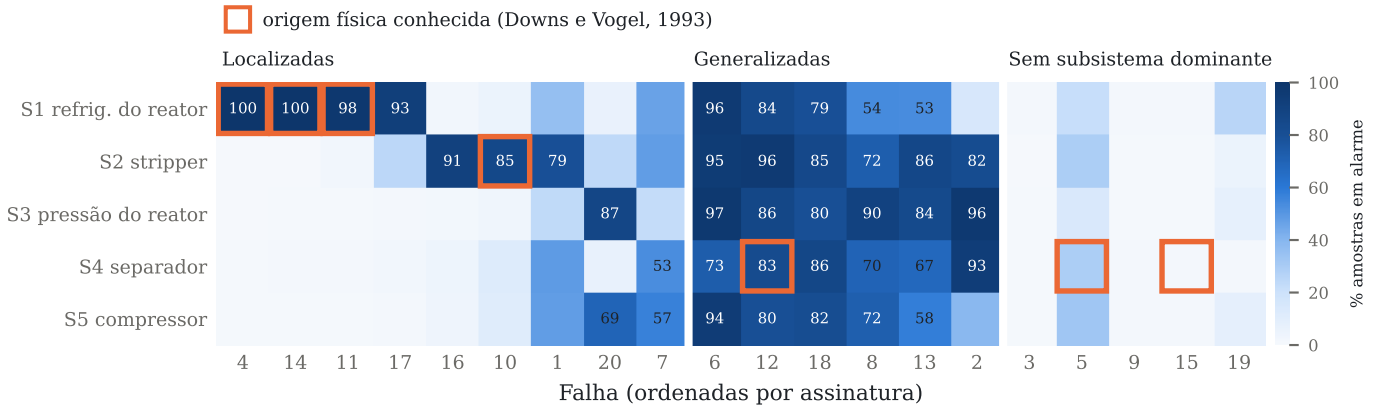


Fig. 6. Porcentagem das amostras em falha em que cada subsistema, isoladamente, excede seu limiar (resíduo de simulação livre do ARX, $W = 10$), com as falhas reordenadas por assinatura: localizadas (até dois subsistemas acima de 50%), generalizadas (três ou mais) e sem subsistema dominante; os valores acima de 50% estão anotados, e o contorno indica o subsistema de origem física conhecida, segundo Downs e Vogel [3].

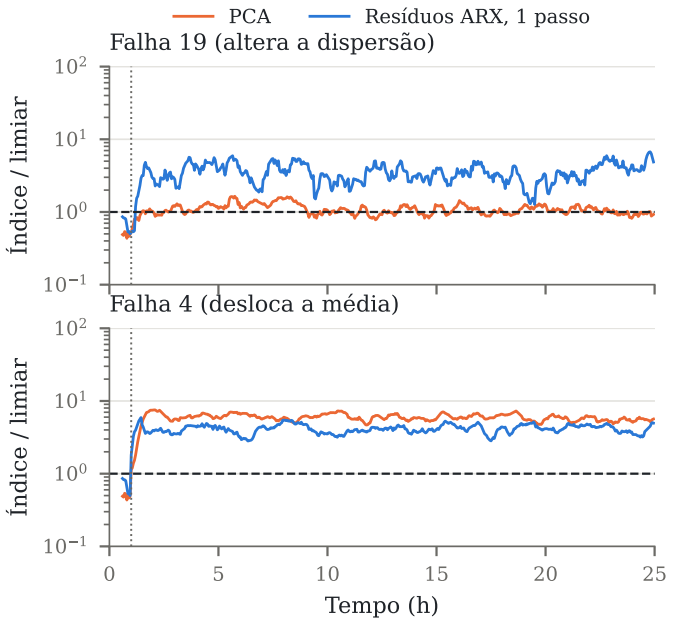


Fig. 7. Índices de detecção normalizados pelo limiar (linha tracejada) na simulação de teste 426, sob as falhas 19 e 4; a linha pontilhada vertical marca a introdução da falha (1 h).

operação normal praticamente coincidem, nenhum limiar as separa, pois baixá-lo eleva a taxa de detecção e a de falsos alarmes na mesma proporção; um detector baseado em dados só detecta falhas que deixam assinatura nos dados, e esse resultado é coerente com a literatura, que considera essas falhas indetectáveis com as medições disponíveis [2], [12]. Quanto aos modelos, os detectores ARX e NARMAX tiveram desempenho praticamente igual, consequência direta de Q1 e Q2, e o NARX neural foi superior ao ARX em resíduos OSA (86,8%), sobretudo na falha 1, em que as variáveis saem da faixa de treino e a saturação da rede possivelmente amplia o erro de predição, efeito que decorreria da extrapolação, e não de uma melhor representação da operação normal, já que o R^2 da rede no teste normal é igual ao do ARX. Por fim, sem a média móvel ($W = 1$), todos os detectores perdem

Tabela X
FALHAS SEM ASSINATURA: NATUREZA SEGUNDO DOWNS E VOGEL [3], MAIOR DESLOCAMENTO DE MÉDIA E MAIOR RAZÃO ENTRE DESVIOS PADRÃO NAS 33 VARIÁVEIS CONTÍNUAS (EDA) E FDR DA FUSÃO

Falha	Natureza	Desloc.	Razão	FDR fusão
3	Degrau na temperatura da alimentação D	$0{,}26\sigma$	1,02	1,5%
9	Variação aleatória da temperatura de D	$0{,}01\sigma$	1,04	1,6%
15	Agarramento da válvula de água do condensador	$0{,}01\sigma$	1,10	2,1%

de 4 a 17 pontos percentuais de FDR, e os detectores que incluem resíduos de simulação livre são os mais robustos a essa escolha, com perdas de 4 a 5 pontos.

C. Isolamento (Q4)

A Fig. 6 mostra, para cada falha, a fração das amostras posteriores à falha em que o resíduo de simulação livre de cada subsistema, isoladamente, excede seu próprio limiar de 99%. A reordenação das colunas por assinatura separa as falhas em três grupos. Nas falhas 4, 11 e 14, todas relacionadas à água de refrigeração do reator (degrau, variação aleatória e agarramento de válvula), apenas S1 alarma, com proporções de 98 a 100%, e a falha 17, de tipo não divulgado, exibe a mesma assinatura, o que sugere origem no mesmo circuito; nas falhas 10 (variação aleatória da temperatura da alimentação de C), em que S2 alarma em 85% das amostras, e 16 (de tipo não divulgado), só S2, o subsistema do *stripper*; e, na falha 20, S3 e S5, os subsistemas da pressão e do compressor. As falhas que perturbam a planta inteira, como a 6 (perda da alimentação de A) e a 12, fazem alarmar todos os subsistemas. Para verificar o isolamento de forma objetiva, contornou-se na Fig. 6 a célula do subsistema de origem das sete falhas cuja localização Downs e Vogel [3] descrevem e que pertencem aos subsistemas modelados. Nas quatro com origem em S1 e S2 (falhas 4, 11, 14 e 10), o subsistema que alarma coincide com a origem. Nas três com origem no circuito de água do condensador, S4, o resultado é misto: a falha 12 alarma S4 em

83% das amostras, mas também todos os demais subsistemas, pois se propaga pela planta; a falha 5 alarma S4 em apenas 29%, porque o controle compensa o degrau; e a falha 15 não deixa assinatura, como discutido na Seção V. Optou-se por não marcar como não aplicáveis as demais células, pois a localização de cinco falhas não é divulgada, as falhas se propagam entre subsistemas e os valores baixos são justamente o que evidencia a especificidade do isolamento. Aceita-se, assim, H4 para as falhas de efeito localizado, e a matriz da Fig. 6 constitui uma assinatura que pode ser usada num classificador de diagnóstico, etapa que não foi abordada neste trabalho.

VI. Publicação e reprodutibilidade

Conforme o plano de gestão de dados, os dados brutos de Rieth et al. não são redistribuídos, mas referenciados pelo DOI, e foram publicados no Zenodo, na comunidade do grupo de Big Data Analytics da Escola Politécnica, sob o DOI 10.5281/zenodo.24532683 [15], os dados derivados necessários para reproduzir os resultados, a saber: os parâmetros de padronização, as estruturas, os parâmetros e as ERR dos modelos (JSON), os resíduos de todas as simulações de validação e teste (Parquet), os limiares e as métricas por falha e detector (CSV), além do código Python que gera cada etapa, desde a leitura dos RData até as tabelas deste artigo, que são escritas diretamente pelos *scripts* a partir dos resultados. O depósito inclui um arquivo README com proveniência e instruções, o dicionário de dados e os metadados de citação no esquema DataCite e no formato CITATION.cff, segundo os princípios FAIR [14], com licença CC BY 4.0 para os dados e MIT para o código. O mesmo registro reúne ainda a versão MATLAB do código, este artigo e o plano de gestão de dados.

A reprodutibilidade foi verificada de forma independente por uma segunda implementação de todo o *pipeline* em MATLAB, que dispensa *toolboxes* e foi executada no GNU Octave a partir dos mesmos arquivos RData: os modelos ARX, NARX e NARMAX resultaram idênticos, termo a termo e com os mesmos coeficientes, assim como as taxas de detecção de todos os detectores que não envolvem a rede neural (PCA, 86,7%; resíduos OSA e de simulação livre, 90,6%; fusão, 93,3%), os testes de correlação e a matriz de isolamento. Apenas os resultados da rede neural diferem, pois dependem da inicialização aleatória, que é distinta nas duas linguagens; o R^2 OSA da rede coincidiu até a terceira casa decimal, e a diferença nas taxas de detecção da rede não altera nenhuma conclusão, já que ela não superou o ARX em operação normal.

VII. Conclusões

O projeto mostrou que modelos de entrada e saída identificados apenas com dados de operação normal fornecem resíduos capazes de detectar a maior parte das falhas do TEP com a mesma taxa de falsos alarmes de um detector multivariado clássico, e de detectá-las melhor quando a falha altera a dinâmica sem deslocar a média, além de indicar o subsistema afetado. A investigação da não linearidade produziu um resultado negativo, mas informativo: nem os termos polinomiais nem a rede neural melhoraram a predição em operação normal, de modo que, nesse conjunto de dados, a estrutura NARX se reduz à ARX, e o ganho de desempenho veio da combinação dos resíduos de predição e de simulação livre e da fusão com a PCA, e não da complexidade do modelo. A Tabela XI resume o resultado de cada hipótese.

Tabela XI
Resultado das hipóteses

H	Resultado	Evidência
H1	Rejeitada	Nenhum termo de grau 2 supera o limiar de ERR; a rede neural empata com o ARX
H2	Aceita em parte	O NARMAX reduz o EQM em 5,3% em S2 e é rejeitado na validação em S3, S4 e S5
H3	Confirmada	FDR de 90,6% contra 86,7% da PCA; ganhos de 18 a 41 pontos nas falhas 10, 16, 19 e 20
H4	Confirmada	As falhas localizadas alarmam apenas o subsistema de origem (4, 11 e 14 em S1; 10 em S2)
H5	Confirmada	As falhas 3, 9 e 15 permanecem no nível da taxa de falsos alarmes

Do ponto de vista da ciência dos dados, a experiência do grupo destacou três aspectos. O primeiro é o volume: os 5 milhões de registros do arquivo de falhas exigiram leitura sequencial e precisão simples para caberem na memória de um computador pessoal, e a modelagem do banco por simulação, com a matriz de cada simulação num único documento, tornou as consultas da análise simples e rápidas. O segundo é o cuidado com a separação dos dados, feita por simulação, e com a calibração dos limiares em dados de validação, sem os quais as taxas de detecção seriam otimistas. O terceiro é o valor da análise exploratória, que antecipou a exclusão das composições, a escolha dos subsistemas, a utilidade da janela móvel e a indetectabilidade das falhas 3, 9 e 15, todas confirmadas na etapa de análise. Como trabalhos futuros, sugerem-se o aumento dos atrasos das entradas em S4 e S5, cujos resíduos ainda se correlacionam com as entradas, a identificação dos modelos em faixas de operação mais amplas, nas quais as não linearidades possam se manifestar, o uso dos arquivos de teste de Rieth et al., com 960 amostras por simulação, e um classificador de diagnóstico baseado nas assinaturas de isolamento.

APÊNDICE

LISTA DE SIGLAS

ARX	Modelo autorregressivo com entradas exógenas
CC BY	Licença Creative Commons de atribuição
CSV	Valores separados por vírgula
DOI	Identificador de objeto digital
EDA	Análise exploratória de dados
EQM	Erro quadrático médio
ERR	Taxa de redução de erro (*error reduction ratio*)
FAIR	Encontrável, acessível, interoperável e reutilizável
FAR	Taxa de falsos alarmes (*false alarm rate*)
FDR	Taxa de detecção de falhas (*fault detection rate*)
FROLS	Mínimos quadrados ortogonais com regressão progressiva (*forward regression orthogonal least squares*)
IC	Intervalo de confiança
JSON	Notação de objetos JavaScript
MLP	Perceptron multicamadas (*multilayer perceptron*)
MIT	Licença de software do Massachusetts Institute of Technology
NARMAX	Modelo NARX com média móvel do ruído
NARX	Modelo ARX não linear
NoSQL	Banco de dados não relacional
OSA	Predição um passo à frente (*one-step-ahead*)
PCA	Análise de componentes principais
R^2	Coeficiente de determinação
SPE	Erro quadrático de predição (*squared prediction error*)
T^2	Estatística T^2 de Hotelling
TEP	Processo Tennessee Eastman
xmeas, xmv	Variáveis medidas e manipuladas do TEP

REFERÊNCIAS

[1] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control*, vol. 36, no. 2, pp. 220–234, 2012.

[2] L. H. Chiang, E. L. Russell e R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London: Springer, 2001.

[3] J. J. Downs e E. F. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Engineering*, vol. 17, no. 3, pp. 245–255, 1993.

[4] C. A. Rieth, B. D. Amsel, R. Tran e M. B. Cook, "Additional Tennessee Eastman process simulation data for anomaly detection evaluation," Harvard Dataverse, V1, 2017, doi: 10.7910/DVN/6C3JR1.

[5] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Chichester: Wiley, 2013.

[6] S. Chen, S. A. Billings e W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[7] W. Michener *et al.*, "Participatory design of DataONE: enabling cyberinfrastructure for the biological and environmental sciences," *Ecological Informatics*, vol. 11, pp. 5–15, 2012.

[8] W. R. Lacerda, L. P. C. Andrade, S. C. P. Oliveira e S. A. M. Martins, "SysIdentPy: a Python package for system identification using NARMAX models," *Journal of Open Source Software*, vol. 5, no. 54, p. 2384, 2020.

[9] F. Pedregosa *et al.*, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[10] B. Efron e R. J. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

[11] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[12] E. L. Russell, L. H. Chiang e R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 1, pp. 81–93, 2000.

[13] T. Munzner, *Visualization Analysis and Design*. Boca Raton: CRC Press (A K Peters Visualization Series), 2014.

[14] M. D. Wilkinson *et al.*, "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, art. 160018, 2016.

[15] H. G. Santana, H. B. Ferreira e G. R. de Souza, "Detecção de falhas no processo Tennessee Eastman por resíduos de modelos NARX/NARMAX: dados derivados, código e documentos," Zenodo, v1.0.0, 2026, doi: 10.5281/zenodo.24532683.