A Machine Learning Pressure Emulator for Hydrogen Embrittlement

Minh Triet Chau [1] **João Lucas de Sousa Almeida** [2] **Elie Alhajjar** [3] **Alberto Costa Nogueira Junior** [2]

Abstract

A recent alternative for hydrogen transportation as a mixture with natural gas is blending it into natural gas pipelines. However, hydrogen embrittlement of material is a major concern for scientists and gas installation designers to avoid process failures. In this paper, we propose a physics-informed machine learning model to predict the gas pressure on the pipes' inner wall. Despite its high-fidelity results, the current PDE-based simulators are time- and computationally-demanding. Using simulation data, we train an ML model to predict the pressure on the pipelines' inner walls, which is a first step for pipeline system surveillance. We found that the physics-based method outperformed the purely data-driven method and satisfy the physical constraints of the gas flow system.

1. Introduction

Climate change requires clean and effective energy storage to replace gasoline, coal, or natural gas (NG). Batteries are a clean carrier but do not have sufficient energy density for sectors such as cement, steel, and long-haul transport (Emma et al., 2021). For those industries, one option that has received considerable attention is low-carbon hydrogen (McQueen et al., 2020), which can store a large amount of energy and does not release greenhouse pollutants in combustion. However, the inefficiency of green H_2 manufacturing process is one of the biggest obstacles to its dissemination (Joshi et al., 2022). While finding an environmentally friendly and affordable way to produce H_2 is a long-term task, it should not deny us hydrogen's immediate benefit.

One viable strategy is blending H_2 with NG (HCNG) to reduce emissions when burning (Melaina et al., 2013). By increasing the volume of H_2 from 0% to 15%, up to 50% reduction in CO_2 emission was observed (Pandey et al., 2022). Blends with less than 20% H_2 by volume can be transmitted by repurposing existing natural gas pipelines, which are 67% cheaper than building new ones (Peter et al., 2020). However, a major drawback with repurposed pipelines is during daily consumption, gas pressure may reach excessive values that lead to hydrogen diffusion through the most current pipeline wall materials (EU Agency for the Cooperation of Energy Regulators, 2021). Specifically, due to friction incurred on the inner wall caused by the gas flow, atomic H can permeate into its metal lattice, reducing the stress required for cracks to form. This phenomenon, known as hydrogen embrittlement (HE), causes pipelines to be prone to leaking H_2, which can lead to catastrophic events due to H_2 ignition in the presence of air, as well as some other complications like decreasing the upper stratospheric ozone mixing ratios (Nicola et al., 2022). Such a risk is currently prohibiting HCNG from becoming more popular. In Germany, where it is most widely adopted, HCNG only accounts for 10% of demand per capita (Dolci et al., 2019).

Preventing HE requires monitoring, operational pressure management, and pipeline maintenance (Ronevich & San Marchi, 2019). To the best of our knowledge, few works frame pipeline monitoring works from a data driven perspective (Spandonidis et al., 2022) while the rest rely on signals from sensors and hardware (Du et al., 2016; Zhu et al., 2017; Aba et al., 2021). To apply ML to this problem, there are two steps involved. The first is addressing the prediction task of the gas flow pressure. The next step is to use the predicted pressure as the input to apply a H_2 diffusion model through the pipe wall (Fick, 1855; Hafsi et al., 2018) to pinpoint in which segment of the pipe the next leakage will be likely to happen. Out of these two issues, we focus on the turbulence gas flow modeling since it is not only a prerequisite but also a more intricate problem than diffusion. In this work, we propose a supervised machine learning-based model[1] that predicts future pressure values from previously observed data. Namely, we implement an Operator Inference prototype for pipe surveillance and contrast it with transformer techniques.

[1]Independent researcher [2]IBM Research Brazil [3]RAND Corporation, USA. Correspondence to: Minh Triet Chau <s6michau@uni-bonn.de>.

Accepted after peer-review at the 1st workshop on Synergy of Scientific and Machine Learning Modeling, SynS & ML ICML, Honolulu, Hawaii, USA. July, 2023.

[1]Source code is available at `https://github.com/minhtriet/hydrogen_emb`.

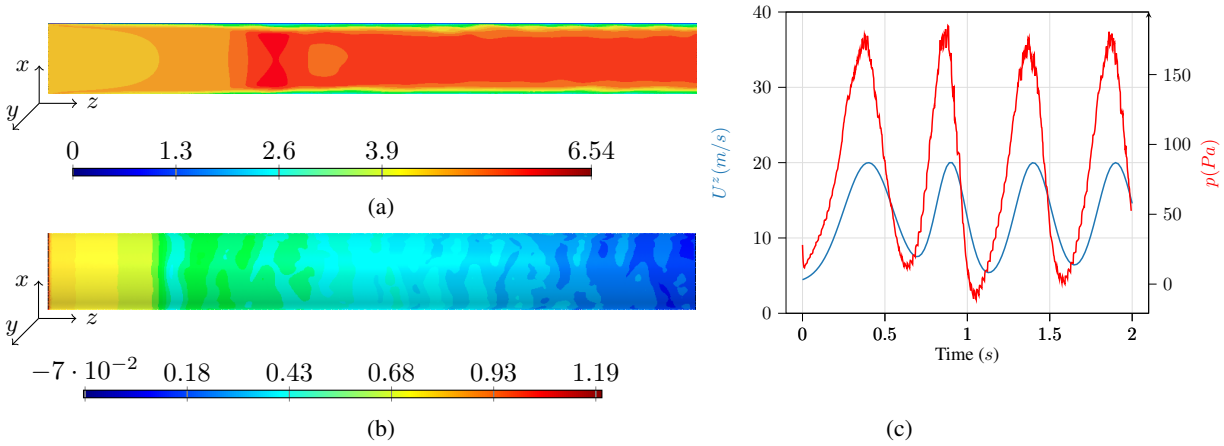

Figure 1. (1a): The velocity field profile $\|\mathbf{U}\|_2$ of gas flow (m/s) in a pipeline. (1b): The pressure field of gas flow to the inner wall. (1c) The inlet velocity U^z and the mean inner wall pressure p of the whole pipeline through time. The periodic profile for U^z is suitable to imitate the real-life demand (Su et al., 2019).

The paper is organized as follows. After this short introduction, we explain the dynamical system and the simulation setup for our experiments in Section 2. In Section 3, we discuss the results of these simulations and how they compare to other baselines. We conclude the paper with limitations to our work and some future research directions.

2. Problem

Physics viewpoint. The problem at hand is modeling a turbulent flow. This is because the Reynolds number of the flow is $Re = \frac{\rho u D}{\mu} \approx 15,565.58$ Here, ρ is the density, u is the flow speed, D is diameter and μ is H_2 dynamic viscosity. When the Reynolds number is larger than 2900, the flow is turbulent (Schlichting & Gersten, 2016), which makes the problem chaotic and tricky for ML models to extrapolate (See Figure 1) . For such systems, the most accurate method, as well as time- and computationally-demanding is the Navier-Stokes equation. For a 2D flow, it is defined as

$$\tau[\frac{\partial U_i}{\partial t} + U_j \frac{\partial U_i}{\partial x_j}] = -\frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}[T_{ij}^{(v)} - \tau\langle u_i u_j\rangle]$$

Here, δ_{ij} is the Kronecker delta, $U_\bullet$ and $u_\bullet$ are the mean and fluctuating velocity, P is the mean static pressure, $T_{ij}^v = 2\mu\tilde{s}_{ij}$ is the viscous stresses, $\tilde{s}_{ij}$ $\tilde{s}_{ij} = \frac{1}{2}[\frac{\partial \tilde{u}_i}{\partial x_j} + \frac{\partial \tilde{u}_j}{\partial x_i}]$ is the instantaneous strain rate tensor, $\tau\langle u_i u_j\rangle$ is the Reynolds shear stress.

Assuming a linear relationship with the mean flow straining field and incompressible flow, $\tau\langle u_i u_j\rangle = -2\mu_t(S_{ij} - \frac{1}{3}S_{kk}\delta_{ij})$. Here $S_{ij} = \frac{1}{2}[\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}]$ is the mean strain rate. To solve it, we use the $k-\omega$ SST model (Menter, 1993), a linear eddy-viscosity model. While being highly accurate, solving it requires tremendous time and computation power.

Table 1. Simulation setup. The specifications of the pipeline, U^z and p_{outlet} are industrial standard from (Mohitpour et al., 2003).

Parameter	Value
Pipeline	
Diameter (cm)	7.62
Length (cm)	500
Boundary condition (BC)	
Inlet velocity U^z (m/s)	Periodic profile in Figure 1c
Relative p_{outlet} (pa)	0
Wall	No slip
Simulation control	
Δt	$0.002s$

Simulation setup. We ran a SimScale simulation with the specifications in Table 1 to generate the raw data. Even with only 1000 time steps with fixed $\Delta t = 0.002s$, it took 8.4 hours to finish.

3. Experiments

Set up For the training data, we select features from the set of $(\mathbf{U}, p, k, \omega, nut)$ and organize them similarly to the same matrix below, with m the temporal and n the spatial dimension. Each algorithm in our experiment will predict the future p, similarly to Figure 1b. Here, $\mathbf{U} \in \mathbb{R}^3$ is the velocity in x, y, z axis, p is the pipeline internal static pressure, k is the rate of dissipation of kinematic turbulence, ω is the kinematic turbulence energy and nut is the kinematic

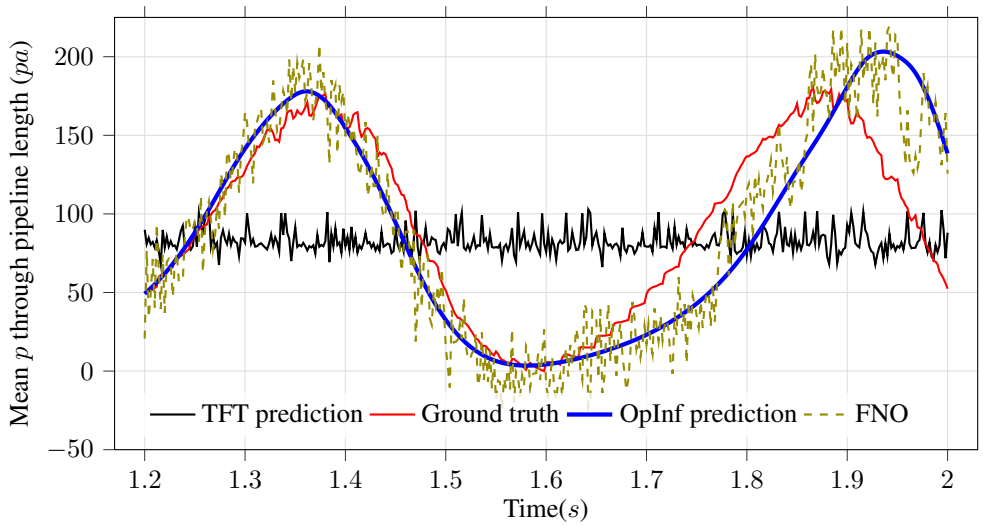

Figure 2. Pressure prediction in the test set.

turbulent viscosity.

$$\begin{pmatrix} k_{1,1} & k_{1,2} & \cdots & k_{1,n} \\ p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ \mathbf{U}_{1,1} & \mathbf{U}_{1,2} & \cdots & \mathbf{U}_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{U}_{m,1} & \mathbf{U}_{m,2} & \cdots & \mathbf{U}_{m,n} \end{pmatrix}$$

The first 50% of data is for the training set, and the next 10% is the validation set. After that, we retrain the model with the train and validation sets to predict the test set. To reduce the computational cost, we first apply PCA to generate a low-rank representation of the dynamic system. Then, we apply Transformers, Temporal Fusion Transformers (TFT) (Lim et al., 2021), Operator Inference (Ghattas & Willcox, 2021) and Fourier Neural Operator (Li et al., 2020). The latter two are physics-aware methods in the reduced order space. Each method will predict 10 time steps ahead, then use its own prediction as the input, similarly to a rolling window. At the end of the validation set, each algorithm will immediately produce a $\mathbb{R}^{m_{test} \times n}$ matrix.

Note that while U^z is available on training, it is absent on testing. Every method has to predict k time-step forward from the end of the training data to the end of testing data without seeing the testing data . This is because the U^z is generally not available at the time of running inference to the future.

Baseline: Transformer and TFT. In this approach, we perform a grid search for the hyperparameters, such as the number of encoder and decoder layers, dropout rate, and the number of attention heads. Except for p, the other features are set as covariance series. The training of those models is powered by (Herzen et al., 2022).

Fourier Neural Operator. FNO aims to learn a neural operator, or neural network, to map from the input to the PDE solution by parameterizing the integral transformation in a Fourier space. The space is chosen to exploit the fact that differentiation is equivalent to multiplication in the Fourier domain.

Operator Inference (OpInf). Operator Inference assumes a low dimensional manifold from the high fidelity model. It reduces a full order model to a polynomial representation in latent space as

$$\frac{d}{dt}x_t = c + A(x_t) + H(x_t \otimes x_t) + G(x_t \otimes x_t \otimes x_t) + B(u_t)$$

Here $\otimes$ operator denotes a column-wise Kronecker product. A, H and G represent linear, quadratic, and cubic state matrices respectively. $B(u_t)$ can be used as the reduced operator of a forcing term in reduced space. c is a constant term. Each of the terms may or may not be used by OpInf, depending on the experimenter's choice. A, H and G are learned by solving a regularized least-squares problem. The nonlinear parts of the full-order model are included since many meaningful PDEs have quadratic nonlinearities in their operators. Combining the linear PCA transformation with the quadratic manifold in latent space is equivalent to applying Galerkin Projection to the original full-order model using a truncated PCA basis. In this work, we use A, H, B operators and c.

Besides model selection, we also have to find the optimal number of the bases for its dimension reduction, and the regularizer value for the least square solver since unregularized RMSE minimization may lead to overfitting. For the

Table 2. Summary for Transformer, TFT, and OpInf approach.

Method	Transformer	TFT	OpInf	FNO
Best input feature	p, U^z	p, U^z	p, U^z	p
RMSE	60.5701	58.3013	**34.7737**	36.8884

number of bases, we measure the quality of the reduced order model by computing the preserved energy with respect to the original dynamical system and calculating the RMSE between the original state variables and a PCA reconstruction of those variables. We found out that regardless of the subsets of input $(p, \mathbf{U}, k, nut, \omega)$, 30 basis capture at least 99.82% of the cumulative energy of the system. We also perform a PCA on the train and test set with the same number of bases to check if there is a distribution shift between them. The RMSE of the models with their best input set are in Table 2. See Figure 2 for a visual comparison.

4. Discussion

The high RMSE of Transformers. At first, it seems counter-intuitive that a successful model like Transformers behaves like a mean estimator in the test set. In fact, (Zeng et al., 2023) suggests that transformers are ineffective for time-series prediction, even when input data series are measured in years, let alone in a couple of thousands time steps in this work. Indeed, for chaotic systems, the required amount of data is tremendous. For example, (Nguyen et al., 2023) uses weather data for ten years to predict reliably for merely one week.

Evidently, in Figure 2, TFT's high bias leads to failure to capture the system dynamics induced by the BC. On the other hand, we experienced the high variance of OpInf method, where it outputs a sinusoidal prediction but its peaks and valleys are wrong. However, decreasing the number of features while keeping the same number of training data took care of the high variance in OpInf. It gives the best result as it uses the two most important features U^z and p. p is what we are trying to predict, and the flow is strictly dominant in the z-axis (See Figure 1). FNO also does a good job. We observe a jagged prediction of neural network methods. Note the typical prediction of Transformers and TFT, regardless of input features and parameter tuning. For OpInf, a higher regularizer causes the insensitivity to changes from BC, even though it minimizes test set loss.

While the simulation is in a short period and small pipeline due to our computational power constraints, it should not hinder its generalization to a bigger pipeline system. It is because the length of the pipe only indicates how pressure drops along the lines, and the pressure equation should follow the same physics equations. For the time-bound question, OpInf has demonstrated its accuracy in complex systems with long prediction window (Sharma & Kramer, 2022) or a high number of snapshots (Swischuk et al., 2020; McQuarrie et al., 2020).

The importance of speeding up with ML. Taking a step back, the HE problem is a control problem, as it concerns what to do when a breach is happening. Therefore, to predict when and where the next leakage might occur, it is best to create a handful of scenarios with different U, p profiles in different pipeline systems. However, turbulence modeling is expensive, even in a pipeline, as noted in Section 2. In comparison, the inference time in ML methods is less than a second compared to 8.4 hours in the numerical simulation, making the incorporation of such ML techniques crucial in future investigations.

The degradation of the prediction performance. In Figure 2, the RMSE is increasing around $t = 1.9s$, which hints at a limitation of the prediction window for physics-aware algorithms. With the current amount of data in our experiments, we cannot quantify the relationship between the size of the dataset and the prediction window length, as well as the confidence interval of the training regime. Both of these pillars will be a part of a follow up paper. Moreover, quantifying how the error in the pressure prediction affects the HE depends on the H_2 diffusion models, which we already discussed in Section 1.

5. Conclusion

In this work, we propose an ML model for emulating HE, which is one of the biggest obstacles against HCNG and H_2 public usage. From an environmental perspective, wider adoption of HCNG can reduce the amount of CO_2 exhaustion. Moreover, preventing H_2 leakage also improves the safety of the transport system and the atmosphere quality of the surrounding area. Compared to numerical simulations, our ML model takes less than a second to produce an accurate inference of the pressure in a pipeline. It is the first step to bring HCNG one step closer to mass adoption today and pure H_2 delivery infrastructure in the future.

Impact statement

This paper lies in the intersection of computational flow dynamics and machine learning. The findings of this pa-

per may have implications beyond the energy sector and encourage interdisciplinary approach in the fields of materials science, physics, and machine learning. When applied, this work should bring hydrogen transportation closer to the reality, which is a positive consequence for the future society.

References

Aba, E. N., Olugboji, O. A., Nasir, A., Olutoye, M. A., and Adedipe, O. Petroleum pipeline monitoring using an internet of things (IoT) platform. *SN Applied Sciences*, 3(2):180, February 2021. ISSN 2523-3963, 2523-3971. doi: 10.1007/s42452-021-04225-z. URL http://link.springer.com/10.1007/s42452-021-04225-z.

Dolci, F., Thomas, D., Hilliard, S., Guerra, C. F., Hancke, R., Ito, H., Jegoux, M., Kreeft, G., Leaver, J., Newborough, M., Proost, J., Robinius, M., Weidner, E., Mansilla, C., and Lucchese, P. Incentives and legal barriers for power-to-hydrogen pathways: An international snapshot. *International Journal of Hydrogen Energy*, 44(23):11394–11401, 2019. ISSN 0360-3199. doi: https://doi.org/10.1016/j.ijhydene.2019.03.045. URL https://www.sciencedirect.com/science/article/pii/S0360319919309693.

Du, G., Huo, L., Kong, Q., and Song, G. Damage detection of pipeline multiple cracks using piezoceramic transducers. *Journal of Vibroengineering*, 18 (5):2828–2838, August 2016. ISSN 1392-8716, 2538-8460. doi: 10.21595/jve.2016.17040. URL https://www.extrica.com/article/17040.

Emma, W., Olivia, H., and Charles, F. Enabling the european hydrogen economy. Technical report, May 2021.

EU Agency for the Cooperation of Energy Regulators. Transporting Pure Hydrogen by Repurposing Existing Gas Infrastructure: Overview of existing studies and reflections on the conditions for repurposing, July 2021. URL https://web.archive.org/web/20220805232342/https://acer.europa.eu/Official_documents/Acts_of_the_Agency/Publication/Transporting%20Pure%20Hydrogen%20by%20Repurposing%20Existing%20Gas%20Infrastructure_Overview%20of%20studies.pdf. publisher: European Union Agency for the Cooperation of Energy Regulators.

Fick, A. Ueber diffusion. *Annalen der Physik und Chemie*, 170(1):59–86, 1855. doi: 10.1002/andp.18551700105. URL https://doi.org/10.1002/andp.18551700105.

Ghattas, O. and Willcox, K. Learning physics-based models from data: perspectives from inverse problems and model reduction. *Acta Numerica*, 30:445–554, 2021. doi: 10.1017/S0962492921000064.

Hafsi, Z., Mishra, M., and Elaoud, S. Hydrogen embrittlement of steel pipelines during transients. *Procedia Structural Integrity*, 13:210–217, 2018. ISSN 2452-3216. doi: https://doi.org/10.1016/j.prostr.2018.12.035. URL https://www.sciencedirect.com/science/article/pii/S2452321618302683. ECF22 - Loading and Environmental effects on Structural Integrity.

Herzen, J., Lässig, F., Piazzetta, S. G., Neuer, T., Tafti, L., Raille, G., Pottelbergh, T. V., Pasieka, M., Skrodzki, A., Huguenin, N., Dumonal, M., Kościsz, J., Bader, D., Gusset, F., Benheddi, M., Williamson, C., Kosinski, M., Petrik, M., and Grosch, G. Darts: User-friendly modern machine learning for time series. *Journal of Machine Learning Research*, 23(124):1–6, 2022. URL http://jmlr.org/papers/v23/21-1177.html.

Joshi, M., Chernyakhovskiy, I., and Chung, M. Hydrogen 101: Frequently asked questions about hydrogen for decarbonization. July 2022. doi: 10.2172/1879231. URL https://www.osti.gov/biblio/1879231.

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for parametric partial differential equations, 2020. URL https://arxiv.org/abs/2010.08895.

Lim, B., Loeff, N., Arik, S., and Pfister, T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. 2021.

McQuarrie, S. A., Huang, C., and Willcox, K. Data-driven reduced-order models via regularized operator inference for a single-injector combustion process. *CoRR*, abs/2008.02862, 2020. URL https://arxiv.org/abs/2008.02862.

McQueen, S., Stanford, J., Satyapal, S., Miller, E., Stetson, N., Papageorgopoulos, D., Rustagi, N., Arjona, V., Adams, J., Randolph, K., Peterson, D., Hill, L., Koleva, M., Reinhardt, T., Frye, E., Schrecengost, R., Kokkinos, A., Litynski, J., Conrad, R., Soloveichik, G., Tew, D., Litzelman, S., Vetrano, J., Onuschak, R., Hahn, A., Hsieh, E., and Costa, R. Department of energy hydrogen program plan. 11 2020. doi: 10.2172/1721803. URL https://www.osti.gov/biblio/1721803.

Melaina, M. W., Antonia, O., and Penev, M. Blending hydrogen into natural gas pipeline networks. a review of key issues. 3 2013. doi: 10.2172/1219920. URL https://www.osti.gov/biblio/1219920.

Menter, F. R. Zonal two equation k-w turbulence models for aerodynamic flows. 1993.

Mohitpour, M., Golshan, H., and Murray, A. *Pipeline Design and Construction*. American Society of Mechanical Engineers, Fairfield, CA, July 2003.

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., and Grover, A. Climax: A foundation model for weather and climate, 2023. URL https://arxiv.org/abs/2301.10343.

Nicola, W., Paul, G., James, K., Alexander, A., and John, P. Atmospheric implications of increased hydrogen use, 2022.

Pandey, V., Guluwadi, S., and Tafesse, G. H. Performance and emission study of low HCNG fuel blend in SI engine with fixed ignition timing. *Cogent Engineering*, 9(1):2010925, December 2022. ISSN 2331-1916. doi: 10.1080/23311916.2021.2010925. URL https://www.tandfonline.com/doi/full/10.1080/23311916.2021.2010925.

Peter, A., Frank, H., Christoph, v. d. B., Stefan, E., and Thomas, T. Hydrogen infrastructure – the pillar of energy transition, 2020.

Ronevich, J. A. and San Marchi, C. W. Hydrogen effects on pipeline steels and blending into natural gas. 10 2019. URL https://www.osti.gov/biblio/1646101.

Schlichting, H. and Gersten, K. *Boundary-layer theory*. Springer, Berlin, Germany, 9 edition, October 2016.

Sharma, H. and Kramer, B. Preserving lagrangian structure in data-driven reduced-order modeling of large-scale dynamical systems, 2022. URL https://arxiv.org/abs/2203.06361.

Spandonidis, C., Theodoropoulos, P., Giannopoulos, F., Galiatsatos, N., and Petsa, A. Evaluation of deep learning approaches for oil and gas pipeline leak detection using wireless sensor networks. *Engineering Applications of Artificial Intelligence*, 113:104890, 2022. ISSN 0952-1976. doi: https://doi.org/10.1016/j.engappai.2022.104890. URL https://www.sciencedirect.com/science/article/pii/S0952197622001294.

Su, H., Zio, E., Zhang, J., Xu, M., Li, X., and Zhang, Z. A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced deep-RNN model. *Energy*, 178: 585–597, July 2019. doi: 10.1016/j.energy.2019.04.167. URL https://doi.org/10.1016/j.energy.2019.04.167.

Swischuk, R., Kramer, B., Huang, C., and Willcox, K. Learning physics-based reduced-order models for a single-injector combustion process. *AIAA Journal*, 58 (6):2658–2672, jun 2020. doi: 10.2514/1.j058943. URL https://doi.org/10.2514%2F1.j058943.

Zeng, A., Chen, M., Zhang, L., and Xu, Q. Are transformers effective for time series forecasting? 2023.

Zhu, J., Ren, L., Ho, S.-C., Jia, Z., and Song, G. Gas pipeline leakage detection based on PZT sensors. *Smart Materials and Structures*, 26(2): 025022, February 2017. ISSN 0964-1726, 1361-665X. doi: 10.1088/1361-665X/26/2/025022. URL https://iopscience.iop.org/article/10.1088/1361-665X/26/2/025022.