

Machine Learning Surrogate for Thermal Expansion Stress Prediction and Optimal Routing Recommendation of Nuclear Piping Using Accumulated Stress Analysis Data

Yeon-ji Song^{a*}

^aAI Convergence Design Team, KEPCO Engineering & Construction Co., Ltd.

*Corresponding author: julia@kepc-enc.com

***Keywords :** piping stress analysis, surrogate model, pipe routing, thermal expansion, machine learning

1. Introduction

In nuclear piping design, thermal expansion stress is evaluated against the ASME Section III NC-3653.2 Eq. (10a) criterion using the piping stress analysis program PIPESTRESS (PEPS) [1,2]. In current practice, an engineer determines a routing configuration from experience and verifies it by analysis; whenever the criterion is violated, the routing is modified and re-analyzed until it passes. Because the number of iterations cannot be predicted in advance, difficult subsystems may require more than ten iterations over up to two weeks, which is a structural cause of calculation-report schedule delay.

Meanwhile, a large archive of past analyses—input files (FRE) defining geometry, restraints and loads, and result files (PRC) containing nodal code-check results—has been kept only for licensing purposes and never reused for new designs. This study converts the archive into a training asset and builds a machine learning surrogate [3] that ranks candidate routing configurations by thermal expansion stress before any analysis is performed.

2. Methods and Results

2.1 Training database construction

A parser was developed to reconstruct three-dimensional piping geometry from FRE files and to join nodal stress ratios (SE/SA) extracted from PRC files as labels. The database covers six subsystems of three material groups (carbon steel SA106-B, alloy steel SA335-P91, stainless steel SA312-TP304), operating temperature differences of 33–185 °C and pipe sizes of DN40–DN200. Shape variants were generated automatically by applying axis-wise scaling factors to the original routing, and each variant was analyzed with PEPS to form pairs that differ only in geometry. The final table contains 201 cases and 14,845 node rows. Data integrity checks were embedded in the pipeline; cross-checking the self-declared maximum stress ratio of each result file against the parsed rows recovered 424 silently dropped nodes (3.0%), most of which were located at stress-concentration points.

2.2 Global span-wise features

Twenty-three local nodal features (bend/valve flags, stress intensification factor, distances to restraints, thermal load index, etc.) were first used for training. Although the local-feature model achieved a high cross-validated accuracy (CV-R² 0.95), a blind re-analysis test on three unseen candidates revealed that the recommended ranking was exactly inverted (Spearman $\rho = -1.00$): the maximum-stress nodes of the candidates had identical local characteristics, so the flexibility difference of the whole route was invisible to the model. Six global span-wise features were therefore introduced, describing the entire flexible span between restraint points: span path length, chord length, tortuosity (path/chord), vertical offset, vertical-component fraction, and thermal-growth absorption margin. With the global features, the reconstruction ranking reversed to $\rho = +1.00$, and a blind test on four additional unseen shapes gave $\rho = +0.80$.

2.3 Model training

XGBoost [4] and Random Forest regressors were compared under case-wise GroupKFold cross-validation so that all nodes of a case are excluded together from training. XGBoost was selected (CV-R² 0.897 ± 0.006 , MAE 0.0297 versus 0.828 for Random Forest). The case-level rank correlation of maximum stress ratio over the 201 cases was 0.938. Feature-importance analysis shows the six global features jointly account for 0.313 of total importance, confirming that they act as a substantive basis of the ranking decision.

2.4 Blind validation by re-analysis

For the final validation, seven new candidate routings connecting the same anchor pair were evaluated under a discriminating condition ($\Delta T = 150$ °C), re-analyzed with PEPS, and scored by leave-one-out after the verified cases were ingested into the training set (active learning). The predicted ranking agreed with the re-analyzed ranking with Spearman $\rho = +0.893$ ($p = 0.007$), and the best configuration was correctly identified (Fig. 1). The trained model was deployed as a web tool that ranks 26 auto-generated candidates within one second from

anchor coordinates and operating conditions only, so that engineers can start from a favorable configuration before any stress analysis.

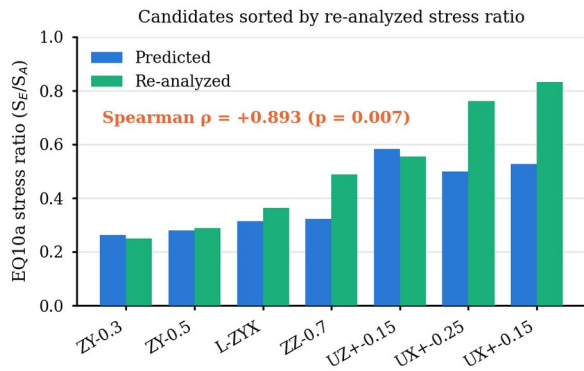


Fig. 1. Predicted vs. re-analyzed stress ratios of seven unseen routing candidates (leave-one-out).

3. Conclusions

Archived piping stress analyses were converted into a reusable training asset, and a surrogate model that ranks routing candidates by thermal expansion stress was developed and validated against actual re-analyses ($\rho = +0.893$). The key enabler was the global span-wise feature set, without which candidate ranking was impossible regardless of model accuracy. The tool is inserted before—not in place of—the licensed analysis program, reducing trial-and-error iterations to one or two confirmation runs. Future work includes extending the database to additional subsystems and dynamic load cases such as seismic loads and migrating to graph neural networks that learn the piping connectivity directly.

REFERENCES

- [1] ASME Boiler and Pressure Vessel Code, Section III, Division 1, Subsection NC, The American Society of Mechanical Engineers.
- [2] PIPESTRESS User's Manual, DST Computer Services S.A., Zurich, Switzerland.
- [3] A. I. J. Forrester, A. Sobester, and A. J. Keane, Engineering Design via Surrogate Modelling: A Practical Guide, Wiley, 2008.
- [4] T. Chen and C. Guestrin, XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD, pp. 785-794, 2016.
- [5] L. Breiman, Random Forests, Machine Learning, Vol. 45, No. 1, pp. 5-32, 2001.