CONFERENCE PRE-PRINT

CATALOGUE-BASED REVERSE ENGINEERING: FOR AI-BASED MODELLING IN FUSION REMOTE MAINTENANCE EQUIPMENT DESIGN

William Brace
VTT Technical Research Centre of Finland Ltd,
Espoo, Finland
Email: william.brace@vtt.com

Van-Dung Truong
VTT Technical Research Centre of Finland Ltd,
Espoo, Finland

Qais Saifi
VTT Technical Research Centre of Finland Ltd
Espoo, Finland

Abstract

Maintainability is a mission-critical aspect of the EUROfusion DEMO reactor design, crucial for ensuring the safety and reliability of complex engineering applications in nuclear reactors, making the development of remote maintenance equipment integral to the process. The design process faces challenges from uncertainty, unstable requirements, and frequent engineering changes in the reactor architecture. In addition, the DEMO design has evolved into a complex process that leverages previous experience and creativity to generate innovative solutions built upon existing products with a higher technology readiness level. This paper presents the application of Artificial Intelligence modelling, aligned with catalogue-based reverse engineering, to reduce uncertainty and complexity, and foster creativity, given the ill-defined requirements and ongoing changes. AI models—built with surrogate machine learning mirror conceptual design steps such as reasoning and problem-solving. Data collection is foundational, involving the integration of commercial-off-the-shelf technologies with higher maturity and technology readiness levels, as well as contextual and reverse engineering analyses to create digital data as a knowledge repository. The methodology progresses through three phases: (1) building a data repository via catalogue-based reverse engineering; (2) contextual definition to analyse and refine information from the data repository to extract high-level data for the data-driven modelling; and (3) modelling and simulating the hybrid AI optimisation, integrating the surrogate models with multiphysics models and an inverse algorithm for optimised design solutions and recommendations. The results from a case study show increased automation and effective solution optimisation. Overall, the AI agent acts as a design assistant, while human decision-making remains central throughout the automated design process.

1. INTRODUCTION

The engineering design of components and systems in a fusion power plant (FPP) includes the development of remote maintenance equipment (RME) to ensure ease of maintenance, operational safety, reliability, and system availability [1]. RME is specially designed for high-risk environments (HRE), enabling safe monitoring and servicing of components without direct contact. HRE equipment requires redesigns with safety features to ensure fail-safe, fault-tolerant operation, maintaining safe performance under extreme conditions where failure could cause environmental damage. Additional designs are incorporated—designs for recoverability and rescuability—to increase reliability and availability, raising the design complexity and technical challenges.

During the FPP design process, several limitations occur due to unstable and insufficient requirements, as well as numerous engineering changes (ECs) in the plant architecture. A novel technology-driven approach, catalogue-based design (CBD), has been introduced to promote technological advancements [2]. The CBD is advocated to incorporate commercial-off-the-shelf (COTS) products, integrating robots with higher maturity and technology readiness levels (TRL); however, they may require redesign to ensure compatibility with the plant architecture. The hypothesis is that implementing data science techniques from machine learning and artificial intelligence (ML/AI) based modeling for RME redesign will reduce design complexity and boost creativity throughout the entire design process. ML/AI are necessary to accelerate progress toward achieving these innovative solutions by maximizing the quantity and utility of data extracted from the CBD.

This paper presents and demonstrates an ML/AI-based model founded on the CBD approach, incorporating the concepts of reverse engineering (RE) and contextual definition (CD) as part of capturing and refining data, and serves as the input block for the ML/AI modeling. The methodology progresses through three phases: (1) building a data repository via catalogue-based reverse engineering; (2) contextual definition to analyse and refine information from the data repository to extract high-level data for the data-driven modelling; and (3) modelling and simulating the hybrid AI optimisation, integrating the surrogate models (SM) with multiphysics models and an inverse algorithm for optimised design solutions and recommendations. The research is ongoing and being applied in the RME development of the DEMO project; this paper outlines the initial outcomes.

2. STATE OF THE ART

2.1. Systems engineering complexity and data extraction

The engineering of FPP components and equipment is rooted in systems engineering complexity (SEC), which encompasses intricate architectures, high-risk environments, and interactions among elements that contribute to the overall function [3]. RME development is within the SEC context, addressing the challenges of designing and interfacing complex systems with high uncertainty. Key principles include understanding system autonomy, designing for safety, configuring RME for fail-safe and fault-tolerant operation, and balancing risks in FPP [4]. Emerging technologies like ML/AI, SM, and simulation-based systems engineering (MBSE) are employed for improved design optimisation [5]. A well-structured process is key to developing an AI model. The data layer, as the foundation, collects and stores high-quality data for data modelling and transformation. One key method is exploratory discursive design (EDD).

EDD is crucial for data collection; however, according to Maher and Poon [6], it is underutilized in the conceptual design phase. Recent studies utilize CBD, a key component of discursive design, to explore COTS systems and their role in FPP [2]. CBD, which involves collecting proven solutions into catalogues, has garnered interest in engineering design research [7]. Digitized CBD serves as a repository of traditional solutions and a knowledge-based system offering a structured data layer and a basis for automated discursive design. Digitized CBD data are compatible with automated knowledge-based processes and can be modeled as class attributes, similar to object-oriented programming (OOP) [8].

In the RME design, CBD data are recorded at an abstract level, and a combined method of RE and CD can be used to extract higher-quality data. RE is an applied science method that enables the collection and archiving of high-quality, measurable data in digital archives. It begins with a finished product, then systematically assesses and evaluates to identify the required data [9]. The CD method is incorporated to identify the set of context variables relevant to supporting the modelling and evaluation of a product. Key methodologies of CD include context analysis, which is applied to extract valuable data from an operational environment, and context variables that are used to discriminate data as constraints or functional requirements [10].

2.2. AI-based modelling

Literature review shows that artificial intelligence (AI) modelling involves complex systems designed to replicate human intelligence or creativity through algorithms and learning mechanisms [11]. Built with methodologies such as machine learning (ML), deep learning, and neural networks, AI models demonstrate cognitive abilities like reasoning, problem-solving, and understanding—key to conceptual design. AI/ML helps manage complexity by automating design processes and exploring options for optimal solutions. In the conceptual design phase, AI/ML supports decision-making by evaluating multiple scenarios in complex systems. Parameter estimation, a core ML process, analyzes data to find optimal solutions and predict models that match data, serving as an inverse problem to determine causes from effects [12]. In fusion research, data-driven ML has made advances in prediction, planning, and generating management systems [13]. This research applies data-driven ML, combining multiphysics modelling with inverse problem-solving and numerical optimization using SM, to develop solutions in the conceptual design of fusion RME.

Contemporary system design involves multiphysics challenges, with nonlinear and unpredictable behaviors from multiple physical phenomena, even if each is linear [14]. This complexity necessitates rigorous analysis, as evident in CAE activities across various industries. Topology optimisation (TO), a subdomain of the inverse problem, utilises the CAE method from mechanics, discretising designs into finite elements (FE), and employs partial differential equation (PDE) constraints, FE analysis, and gradient algorithms to optimise material distribution for goals such as thermal resistance or stress limits [14]. Models are solved via FE, electromagnetic (EM), or Computational Fluid Dynamics (CFD). In fusion, multiphysics issues are critical due to conditions such as radiation, thermal transients, and magnetic fields, which necessitate the use of TO. Examples include adjointbased and TO for divertor and plasma-facing components [15] or neural networks with NSGA-III for blanket design [16]. Formulating design as multiphysics is vital for developing complex fusion RME.

2.2.1. Overview of inverse problems and numerical optimisation

Inverse problems are widely used in physics data measurement, where they infer causes from effects, unlike forward problems, which predict effects from causes [16]. In design, they identify variables to meet performance goals. Mathematically, denote x as a set of inputs (specifications), and y as design performance (outputs). Paolo [17] explained that if a forward function A, maps inputs to outputs, such that:

$$y = A(x) \tag{1}$$

Equation (1) has a unique solution. Solving the inverse problem corresponds to inverting this relationship (assumption that A is invertible):

$$A^{-1}(y) \tag{2}$$

Inverse problems are challenging due to their ill-posed nature, often lacking existence, uniqueness, and stability, unlike well-posed problems. This means a solution may not exist, be unique, or that minor data changes can cause significant variations. Solving inverse problems can be derived as an optimisation corresponding to finding inputs x in a feasible region $x \in \Omega \subseteq R^{nv}$, that minimise $f(x)$ - known as the objective (cost, design criterion) function, as follows [17]:

$$inf_x \; f(x) \tag{3}$$

$f(x)$ depends on the specific physics problem, which is unknown analytically. Solving Equation (3) typically begins with an initial guess x_0: $x_0 \in \Omega \subseteq R^{nv}$, estimating the output y_k, then updating x(k+1) based on x_k using numerical optimisation (NO) algorithms. The process stops when the convergence criterion is met, depending on the chosen NO method. Standard methods include gradient-based algorithms and derivative-free or metaheuristic techniques [18]. Gradient-based methods are most common [19], especially when $f(x)$ is smooth and differentiable. They use the gradient $\nabla f(x)$ or Hessian $\mathrm{H}(x)$ of $f(x)$ to determine the steepest descent direction, guiding the search for the optimal x. Using inverse problems in design optimization is challenging due to high computational costs [18]. Initial forward estimations depend on complex simulations (CFD, FEM), making each iteration resource-intensive, time-consuming, and challenging for inverse problems.

2.2.2. The role of the surrogate model in inverse problem solving and machine learning compatibility

To reduce the computational cost of forward estimation in the inverse problem, surrogate-based optimisation is adopted, replacing high-fidelity simulation with a low-cost SM. A SM is a mathematical substitution model or black-boxing of the original, mapping input-output relations from finite samples of simulation or physical experiments [20]. SMs trade accuracy for speed: they may have limited extrapolation and sensitivity to noise, but provide faster computation, replace complex solvers, and shorten design cycles in physics-based optimisation. Instead of exact simulation, they approximate outputs, accelerating convergence to optimal solutions. Common surrogate types include radial basis function (RBF), Gaussian process (Kriging), linear, and support vector regression [21]. Recent advances in supervised ML, including deep neural networks and gradient-boosted trees, have narrowed the accuracy gap in AI. These methods improve generalization and reduce errors in nonlinear problems [20]. Consequently, SMs can now serve as gradient approximators in gradient-based NO methods and flexible substitutes for costly objective functions, thereby broadening their application in inverse design problems.

3. AI-BASED MODELLING METHODOLOGY AND CASE EXAMPLE

3.1. Methodology

The AI modeling begins with data extraction from COTS products using CBD. Next, this data is contextualized through CD via a context analysis of the FPP's architecture and environment, and the high-level data is discriminated as context variables. Therefore, CD involves risk assessment of using COTS in the FPP, considering uncertainties in design, requirements, and deployment operations. These uncertainties are modeled from three perspectives: plant design assessment using grounded theory data analysis to evaluate COTS technologies;

technology readiness assessment via scatter plot analysis and TRL reevaluations based on the FPP environment; and space reservation assessment for potential incompatibility due to space constraints. These phases are in implementation as part of the ongoing Eurofusion DEMO FP9 project, which aims to integrate COTS as RM equipment into the FPP, with several research outputs published (e.g., [2][4]).

The contextual variable data are input into the hybrid AI-optimization, which comprises a multiphysics modeler, SM, and an inverse problem algorithm. The user may input additional data that affects performance, and the design is initially modeled as a multiphysics problem. The model is simulated to produce limited training data for SM, to act as a substitute for the forward process of the original multiphysics model. This SM enables inverse algorithms to identify optimal design variables for multi-objective optimization, while also serving as a quick method to verify existing design specifications. The inverse algorithm results include optimal design parameters that meet multi-objective goals and validate concepts from the SM. The designer evaluates the results and refines iteratively until an optimal design is achieved; its specifications are fed back into the SM, creating a closed-loop process. The detailed workflow, along with its three phases, is illustrated in *FIG. 1*.

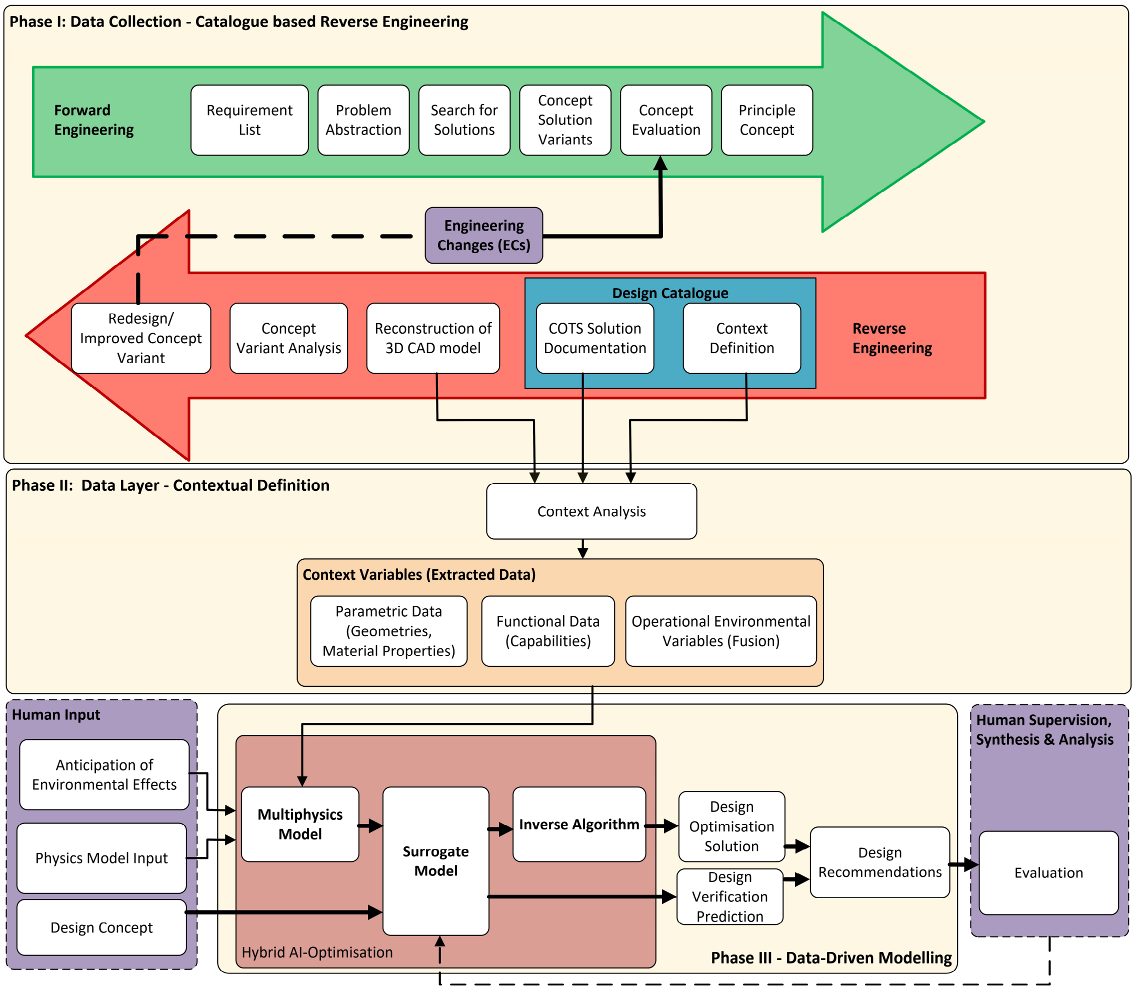

FIG. 1. AI-based modelling methodology for fusion engineering

3.2. Case example

A case study shows how the methodology helps redesign a pipe-climbing robot for welding and inspection in FPP by evaluating COTS robots to operate on various pipe sizes. Using these robots benefits applications in confined spaces, such as in the DEMO piping forest. Results from the CBD identify BOT-WTP27-121 from Beijing Bo Tsing Tech Co., Ltd. as the leading COTS robot for external pipe welding. Identified parametric and functional specifications data were used as the starting reference for the redesign efforts, and the summarized results are shown in TABLE 1.

TABLE 1. KEY ROBOT SPECIFICATIONS AS PARAMETRIC AND FUNCTIONAL DATA

Datasheet	Specifications	Value
Parametric information	Width (mm)	225
	Length (mm)	560
	Height (mm)	480
	Weight (kg)	17
Functional (capability) information	Payload (kg)	60
	Pipe diameter (mm)	168 – 1800
	Pipe thickness (mm)	6 - 60

Given that the COTS robot is engineered for pipe diameters spanning from 168 to 1800 mm and thicknesses ranging from 6 to 60 mm, the pipe dimension data is acquired within this scope in accordance with the ASME B36.10M/B36.19M standards. In the CD assessment, the pipe is modeled as a supported beam made of stainless steel SS316L (N), with support spacing determined based on ASME B31.1/B31.3 guidelines. For each pipe diameter, thicknesses are selected from schedules SCH100, SCHXS, SCH80, and SCH60. TABLE 2 summarizes the considered pipe diameters and corresponding support lengths as part of the parametric dataset.

TABLE 2. PIPE DIMENSIONS AND SUPPORTED LENGTH FOR EACH PIPE SCHEDULE

Pipe	Supported Length
DN200 – DN250	5.8
DN300 – DN350	7
DN400 – DN450	8.2
DN500 – DN550	9.1
DN600	9.8

During maintenance, an emergency pipe flush is required, leading to a rapid decrease in pipe wall temperature. This results in a thermal transient characterized by a steep gradient over a short period. The multiphysics model also represents an extreme scenario where the robot halts at the mid-span of the pipe. Consequently, the model encompasses: the mechanical load due to the bending moment induced by the robot at the mid-span, combined with the pipe's self-weight and water load, as well as the transient thermal stresses caused by the temperature decrease, as shown in *FIG. 2*.

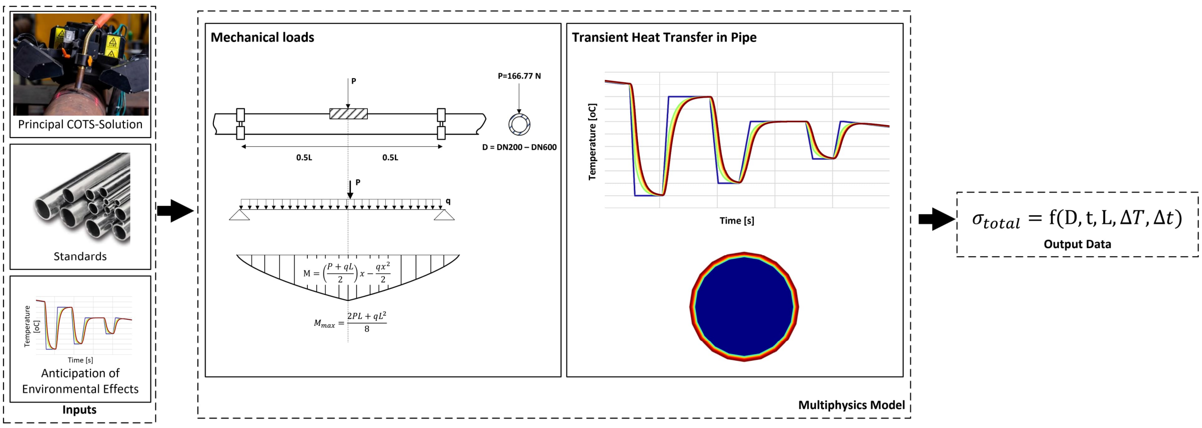

FIG. 2. Descriptions of inputs and the multiphysics model

The design problem input comprises five distinct parameters: pipe diameter D (mm), pipe thickness t (mm), pipe support length L (mm), temperature drop ΔT $(°C)$, and temperature drop duration Δt (s) and one output: the total stress on the pipe at the robot's location $\sigma_{total}(MPa)$. The robot's weight is considered a constant, while D, t and L variables are varied according to pipe standards to represent the possible engineering changes in pipe design dimensions. The simulation test cases were generated at four temperature drops (ΔT = 50, 100, 150, and 200 °C). For each ΔT, the duration of the temperature drop (Δt) was varied across 13 values ranging from 80s to 1040s in increments of 80s. In total, 1820 simulation results were obtained from the full permutation of the varying specified inputs. Multiphysics simulations were conducted to generate limited training data for the SM.

TABLE 3. TRAINING HYPERPARAMETERS

Parameter	**Value**
Batch size	32
Learning rate	0.001
Optimiser	Adam
Epochs	600
Weight decay	0.0001
Loss	Mean-squared error (MSE)

The SM used in this case is Neural Networks, with the hyperparameters shown in TABLE 3, consisting of an input layer of 5 nodes, four hidden layers with 64-64-32-16 nodes respectively, and an output layer of 1 node. The hyperparameters were selected via 5-fold cross-validation on 80% of the physics simulation dataset samples. The final network was evaluated on the remaining 20% samples. The training performance is presented in *FIG. 3* (left), while the prediction performance is shown in *FIG. 3* (right), yielding satisfactory results in different metrics: MSE, RMSE, MAE, and R^2. Specifically, MSE = 2.080 (RMSE = 1.442), MAE = 1.003, R^2= 0.99948.

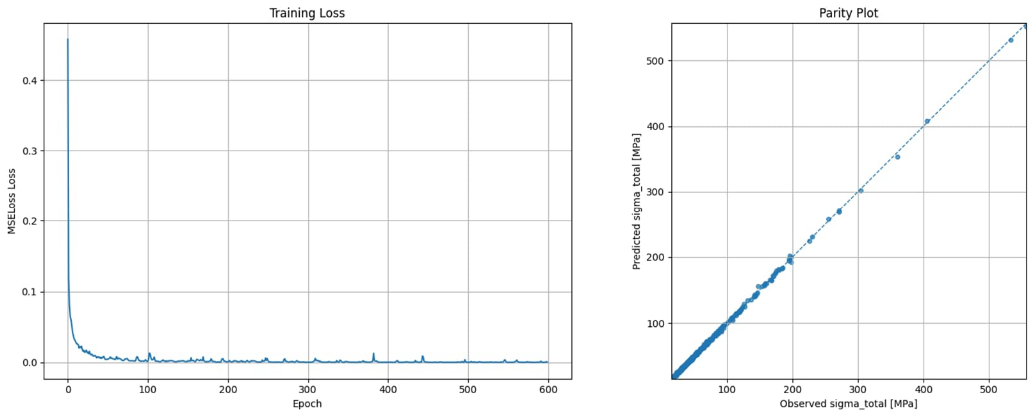

FIG. 3. Loss convergence during data training (left) and parity plot for prediction validation set (right)

The objective of the SM is to support the inverse algorithm and serve as a rapid predictor for design verification. In this context, the principal solution (BOT-WTP27-121) functions as an initial design concept that requires optimization. Utilizing the trained data, the SM predicts continuous scenarios of ΔT and Δt for each standard pipe input, to ascertain whether the current robotic design specifications (mass) would result in permanent pipe deformation due to the cumulative stresses inherent in the multiphysics analysis. Verification is performed through predictions across various pipe schedules and within a specific pipe schedule (*FIG. 4* and *FIG. 5*).

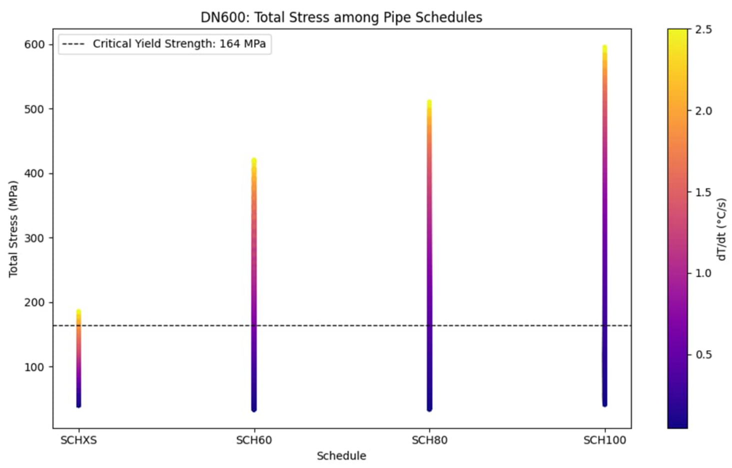

FIG. 4. Prediction for different thermal transient scenarios for the same DN600 pipe among different schedules (thickness)

FIG. 4 shows the prediction surface for the same pipe diameter (DN600) across different pipe schedules. The heatmap bar displays various thermal transient scenarios, with the output indicating the pipe's total stress at the mid-point location. A critical yield strength of 164 MPa for SS316L (N) serves as a reference for the assumed 300 °C peak temperature, representing a conservative design condition that may lead to permanent deformation. Results show that, under current robot specifications, pipe deformations mainly occur in thicker pipes (higher schedule) and worsen with increasing pipe size. This observation is also evident for various pipe dimensions in the same schedule. For instance, *FIG. 5* indicates that among the SCH80 pipes, those with larger diameters and thickness deform more often under various thermal transient scenarios and experience higher stresses.

These results contradict the initial assumption that pipes with higher schedules (and therefore greater thickness) would be more reliable than thinner pipes. This outcome can be attributed to thermal transient phenomena that impose larger temperature gradients and stress distributions across thicker pipe walls, rendering them more susceptible to permanent deformation.

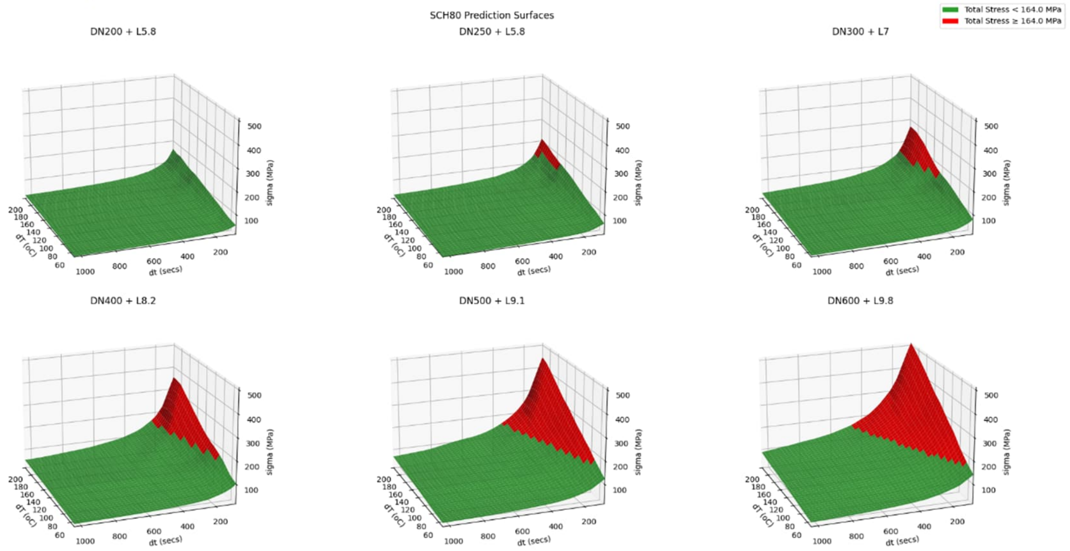

FIG. 5. Prediction surface for different thermal transient scenarios for a different set of design pipe configurations (SCH80)

From a system design perspective, the use of an SM in Phase III enables rapid design verification during the early stages, based on multiphysics simulation data. The prediction results verified that the current robot design specifications are only well-suited for medium and small pipes, where the maximum thermal transient scenarios are less than 1.5 dT/dt to avoid pipe deformation. The majority of pipe failures under intense thermal transient factors are mainly driven by pipe diameter, thickness, and support distance. Consequently, the findings define restrictions on the robot's operational applicability to specific pipe configurations for future operations.

4. CONCLUSIONS

This ongoing research aims to implement robotics and autonomous technologies with high TRL as ready RM equipment for fusion power plant maintenance. It involves AI modeling and simulation to lower redesign complexity caused by high-risk factors and extreme conditions. The paper highlights the first research output: (1) using surrogate models as a design verification tool to replace full Multiphysics simulations, (2) evaluating the impact of the current design on various plant component designs. The research demonstrates that ML/AI modeling can be effectively integrated with a catalogue-based reverse engineering design methodology, which reduces design complexity and enhances creativity, thereby corroborating the research hypothesis. Future work is focused on simulating the inverse algorithm of the hybrid AI-optimizer, utilizing the surrogate model to assist in generating optimized design suggestions. Additional efforts include automating the entire modeling process.

ACKNOWLEDGEMENTS

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

REFERENCES

[1] S. Ciattaglia, G. Federici, L. Barucca, R. Stieglitz, and N. Taylor, "EU DEMO safety and balance of plant design and operating requirements. Issues and possible solutions," Fusion Engineering and Design, vol. 146, pp. 2184–2188, Sep. 2019, doi:10.1016/j.fusengdes.2019.03.149

[2] V.-D. Truong and W. Brace, "Design by cataloguing – a product-driven design approach for the conceptual design of fusion remote maintenance systems," Fusion Engineering and Design, vol. 212, p. 114848, Mar. 2025, doi: 10.1016/j.fusengdes.2025.114848.

[3] Mo Jamshidi, System of Systems Engineering. Wiley, 2008. doi: 10.1002/9780470403501

[4] V.-D. Truong and W. Brace, "Balanced-risk analysis in the engineering design of complex systems with extreme conditions," Fusion Engineering and Design, vol. 208, p. 114690, Nov. 2024, doi: 10.1016/j.fusengdes.2024.114690

[5] D. Humphreys et al., "Advancing Fusion with Machine Learning Research Needs Workshop Report," Journal of Fusion Energy, vol. 39, no. 4, pp. 123–155, Aug. 2020, doi: 10.1007/s10894-020-00258-1.

[6] M. Lou Maher, J. Poon, and S. Boulanger, "Formalising Design Exploration as Co-Evolution," in Advances in Formal Design Methods for CAD, Boston, MA: Springer US, 1996, pp. 3–30. doi: 10.1007/978-0-387-34925-1_1.

[7] K. Roth, "Design catalogues and their usage," in Engineering Design Synthesis, London: Springer London, 2002, pp. 121–129. doi: 10.1007/978-1-4471-3717-7_8.

[8] S. Carlson-Skalak, M. D. White, and Y. Teng, "Using an evolutionary algorithm for catalog design," Res Eng Des, vol. 10, no. 2, pp. 63–83, Jun. 1998, doi: 10.1007/BF01616688.

[9] K. N. Otto and K. L. Wood, "Product Evolution: A Reverse Engineering and Redesign Methodology," Res Eng Des, vol. 10, no. 4, pp. 226–243, Dec. 1998, doi: 10.1007/s001639870003.

[10] Sergio F. Ochoa, lcides Quispe, Julio Ariel Hurtado, and Cecilia Bastarrica, "Project Context Definition Method," 2012.

[11] Oladele Junior Adeyeye and Ibrahim Akanbi, "Artificial Intelligence For Systems Engineering Complexity: A Review On The Use Of Ai And Machine Learning Algorithms," Computer Science & IT Research Journal, vol. 5, no. 4, pp. 787–808, Apr. 2024, doi: 10.51594/csitrj.v5i4.1026.

[12] Richard C. Aster, Brian Borchers, and Clifford H. Thurber, Parameter Estimation and Inverse Problems, 3rd ed. Elsevier, 2019. doi: 10.1016/C2015-0-02458-3.

[13] D. Humphreys et al., "Advancing Fusion with Machine Learning Research Needs Workshop Report," Journal of Fusion Energy, vol. 39, no. 4, pp. 123–155, Aug. 2020, doi: 10.1007/s10894-020-00258-1.

[14] S. Das and A. Sutradhar, "Multi-physics topology optimisation of functionally graded controllable porous structures: Application to heat dissipating problems," Mater Des, vol. 193, p. 108775, Aug. 2020, doi: 10.1016/j.matdes.2020.108775.

[15] B. Curzadd, A. von Müller, R. Neu, and U. von Toussaint, "Topology optimisation of tungsten/copper structures for plasma-facing component applications," Nuclear Fusion, vol. 59, no. 8, p. 086003, Aug. 2019, doi: 10.1088/1741-4326/ab1ff5.

[16] W. Zhou et al., "An intelligent optimisation method for the HCSB blanket based on an improved multi-objective NSGA-III algorithm and an adaptive BP neural network," Nuclear Engineering and Technology, vol. 55, no. 9, pp. 3150–3163, Sep. 2023, doi: 10.1016/j.net.2023.05.024.

[17] P. Di Barba, Multi-objective Shape Design in Electricity and Magnetism, vol. 47. Dordrecht: Springer Netherlands, 2010. doi: 10.1007/978-90-481-3080-1.

[18] S. Koziel and L. Leifsson, Simulation-Driven Design by Knowledge-Based Response Correction Techniques. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-30115-0.

[19] A. R. Conn, K. Scheinberg, and L. N. Vicente, Introduction to Derivative-Free Optimization. Society for Industrial and Applied Mathematics, 2009. doi: 10.1137/1.9780898718768.

[20] D. Šturek and S. Lazarova-Molnar, "Surrogate Modeling: Review and Opportunities for Expert Knowledge Integration," Procedia Comput Sci, vol. 257, pp. 826–833, 2025, doi: 10.1016/j.procs.2025.03.106.

[21] A. Bhosekar and M. Ierapetritou, "Advances in surrogate based modeling, feasibility analysis, and optimisation: A review," Comput Chem Eng, vol. 108, pp. 250–267, Jan. 2018, doi: 10.1016/j.compchemeng.2017.09.017.