POLITECNICO DI TORINO
Repository ISTITUZIONALE

Model order reduction for fusion components: Dynamic mode decomposition model of the magnetron injection gun

Original
Model order reduction for fusion components: Dynamic mode decomposition model of the magnetron injection gun / Novarese, E., Introini, C., Collaku, A., Savoldi, L., Cammi, A.. - In: FUSION ENGINEERING AND DESIGN. - ISSN 0920-3796. - ELETTRONICO. - 232:(2026). [10.1016/j.fusengdes.2026.115974]

Availability:
This version is available at: 11583/3015149 since: 2026-09-03T12:54:04Z

Publisher:
Elsevier Ltd

Published
DOI:10.1016/j.fusengdes.2026.115974

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

07 November 2026

Fusion Engineering and Design 232 (2026) 115974

Contents lists available at ScienceDirect

Fusion Engineering and Design

journal homepage: www.elsevier.com/locate/fusengdes

Model order reduction for fusion components: Dynamic mode decomposition model of the magnetron injection gun ☆

Elia Novarese [a], Carolina Introini [b], Aldo Collaku [a], Laura Savoldi [a], Antonio Cammi [c,b,*]

[a] MAHTEP Group, Dipartimento Energia, Politecnico di Torino, Torino, 10124, Italy
[b] Department of Energy, Politecnico di Milano, Milano, 20156, Italy
[c] Nuclear Technology Center (ENTC), Department of Mechanical and Nuclear Engineering, Khalifa University, Abu Dhabi, 127788, United Arab Emirates

ARTICLE INFO

Keywords:
Model order reduction
Dynamic mode decomposition
Gyrotron
Digital twin
Magnetron injection gun

ABSTRACT

In nuclear fusion systems, the gyrotron is the source of electromagnetic waves, exploiting the electron maser instability to initialize and stabilize the fusion reaction. A simulator that predicts the system's behaviour based on the fundamental physics is crucial for its control. This requires modelling each component by integrating the relevant physical phenomena for that component, and linking the resulting component models together. However, current models of fusion system components like the breeding blanket or the gyrotron are too computationally intensive for real-time simulations if high accuracy is sought and therefore high-fidelity and high-dimensional models are needed, as lumped approaches may not offer high enough accuracy for safety-critical applications such as monitoring and control. Reduced Order Models (ROM) can overcome this problem while maintaining accuracy. This paper focuses on the Magnetron Injection Gun (MIG), responsible for generating the electron beam by thermionic effect. It is essential to calculate the heat transfer within the MIG to predict the electron density on the emitter surface. This methodological paper studies the suitability of Dynamic Mode Decomposition (DMD) in capturing the non-linear dynamics and time-scales of the beam current emission considering a representative pulse provided by the FALCON facility. Results show the computational efficiency and capability of such technique in capturing temporal dynamics, performing spectral analysis, denoising data and forecasting features beyond the training dataset, positioning this paper as a foundation from which more complex models and ROM techniques can be considered.

1. Introduction

For the successful development of nuclear fusion reactors, such as ITER, a key component is the high-performance gyrotron. Gyrotrons are the primary radio-frequency (RF) source for Electron Cyclotron Resonance Heating (ECRH) and current drive, achieving and maintaining the plasma conditions required for fusion.

High-power RF is generated through the maser instability of a high-quality[1] electron beam. This beam is generated by the magnetron injection gun (MIG) from a ring-shaped cathode (the emitter) through thermionic emission, a strongly temperature-dependent phenomenon. To ensure the beam stability and protect the device, it is crucial to accurately model its electrical and thermal dynamics.

High-fidelity models based on 3D geometry are required for simulating electrical and thermal behaviour due to the lack of temperature measures in the MIG. However, these models are unsuitable for control-oriented applications in a digital twin framework because of the complex geometry and broad range of temporal dynamics, where accuracy must be accompanied by quasi-real-time operational times [1]. Broadly speaking, a digital twin is a virtual replica of the physical system that is supposed to run in parallel with it, receiving real-time sensor data as inputs to enable monitoring, prediction and control, hence it must have high accuracy but also quasi-real time running times [2]. To overcome this problem, lumped models can be developed, where spatial dependency of quantities is replaced by their average value. This approximation is valid for components where internal heat conduction is negligible compared to heat transfer with other components. The dimensionless Biot number describes the relative importance of internal conduction (k) compared to heat transfer at external surfaces (h),

☆ This article is part of a Special issue entitled: 'ISFNT-16' published in Fusion Engineering and Design.

* Corresponding author at: Department of Energy, Politecnico di Milano, Milano, 20156, Italy.
E-mail address: antonio.cammi@polimi.it (A. Cammi).

[1] A high-quality beam is a beam with the correct geometrical displacement to excite the correct cavity mode, and with the desired kinetic energy and pitch factor to guarantee proper energy exchange with the electromagnetic field.

https://doi.org/10.1016/j.fusengdes.2026.115974
Received 31 December 2025; Received in revised form 21 July 2026; Accepted 22 July 2026
Available online 8 August 2026
0920-3796/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

according to the formula:

$$\mathrm{Bi} = \frac{h}{k} L,$$

being L the characteristic length of the component Since the key MIG components yields very small Biot numbers (Bi$\ll$ 1), the development of lumped models [3,4] is justified.

However, lumped approaches lack spatial distribution. Model Order Reduction (MOR) techniques strike a balance between the computational efficiency of lumped models and the accuracy of high-fidelity models. MOR aims to extract essential temporal, spatial, and parametric dynamics from a high-fidelity dataset to build low-dimensional surrogate models. Once properly trained, the resulting surrogate model can be used in 'testing' mode, receiving previously-unseen inputs (such as data from the physical system) to estimate the corresponding output faster compared to running the high-fidelity model but with high enough accuracy. Prediction involves estimating parametric transients within the training dataset limits, while forecasting involves estimating the future state beyond the temporal boundaries. Many MOR techniques are widely used in engineering applications (including nuclear fission) [5,6].

Among all available MOR and Machine Learning techniques, this paper uses Dynamic Mode Decomposition (DMD) for several key reasons: (i) DMD computes a low-rank linear operator that is a projection of the true linear system operator, thus providing a direct, data-driven approximation of the underlying system dynamics without requiring the explicit governing equations; (ii) unlike many black-box ML techniques, DMD is mathematically rooted in Koopman theory [7], and the eigenvalues of the low-dimensional surrogate operator are strongly connected to those of the true system operator; (iii) DMD simplifies complex non-linear dynamics into a linear evolution, by approximating the temporal behaviour with a sum of damped or oscillatory exponential modes, making highly non-linear transients easier to forecast; (iv) each DMD mode is associated with a specific eigenvalue that encodes physical properties, allowing for interpretability; (v) variant DMD algorithms, such as the one used in this paper (Optimized DMD), offers robustness against noise and can deal with non-uniformly sampled trajectories.

In light of the above, this paper presents a novel application of MOR techniques for fusion by applying DMD to decompose complex temporal data into spatio-temporal modes [8]. To evaluate how DMD captures nonlinearities and timescales inherent to this problem, the training dataset comes from a validated lumped model of the MIG. This first assessment step of applying the selected ROM algorithm to simple models such as lumped ones is used to verify the suitability of DMD for the particular case of the MIG dynamics: indeed, this work will demonstrates the effectiveness of DMD for this class of problem, providing a foundation for switching to high-fidelity models currently under development.

The structure of the paper is as follows: Section 2 details the mathematical framework for the surrogate model, focusing on DMD, its Optimized variant, and Hankel embedding. Section 3 describes the MIG, followed by the lumped model and validation process in Section 4. Results are presented in Section 5, showing DMD-based surrogate models for latent dynamics identification, spectral analysis, and state forecast. Finally, Section 6 reports key findings and future perspectives.

2. Methodology

2.1. Dynamic mode decomposition

The Dynamic Mode Decomposition (DMD) algorithm was formulated in [8] to discriminate coherent structures in highly-turbulent flows. It is based on the Singular Value Decomposition (SVD) method [5], which decomposes a generic matrix $\mathcal{X}$ as follows:

$$\mathcal{X} = \mathcal{U} \Sigma \mathcal{V}^T, \tag{1}$$

where $\mathcal{U}$ contains the dominant spatial structures or modes (left singular vectors), Σ is a diagonal matrix containing the singular values associated to the left singular vectors in $\mathcal{U}$ and $\mathcal{V}$ contains the temporal dynamics (right singular vectors). The SVD sorts the modes in descending order based on how much information of the original dataset they retain [6], and this information is contained in the singular values: SVD matrices can then be truncated up to a certain rank r, representing the percentage of information retained, and a reduced representation of the original dataset $\mathcal{X}$ can be retrieved:

$$\widetilde{\mathcal{X}} = \widetilde{\mathcal{U}} \widetilde{\Sigma} \widetilde{\mathcal{V}}^T, \tag{2}$$

with $\widetilde{\mathcal{U}} \in \mathbb{R}^{\mathcal{N}_h \times r}$, $\widetilde{\Sigma} \in \mathbb{R}^{r \times r}$ and $\widetilde{\mathcal{V}} \in \mathbb{R}^{\mathcal{N}_t \times r}$.

More in detail, for DMD the snapshot matrix $\mathcal{X}$ is:

$$\mathcal{X} = [x(t_1), x(t_2), \dots, x(t_{\mathcal{N}_t})] \in \mathbb{R}^{\mathcal{N}_h \times \mathcal{N}_t} \tag{3}$$

where $\mathcal{N}_h$ is the number of spatial degrees of freedom (or the number of features) and $\mathcal{N}_t$ is the number of time instants. DMD computes the best-fit linear operator $\mathcal{A}$ that best represents the dataset $\mathcal{X}$, advancing the system in time:

$$\mathbf{x}_{k+1} \approx \mathcal{A} \mathbf{x}_k. \tag{4}$$

The operator $\mathcal{A}$ is obtained from the following optimization problem [9]:

$$\mathcal{A} = \arg \min_{\mathcal{A}^* \in \mathbb{R}^{\mathcal{N}_h \times \mathcal{N}_h}} ||\mathcal{X}^+ - \mathcal{A}^* \mathcal{X}^-||_F = \mathcal{X}^+ (\mathcal{X}^-)^\dagger, \tag{5}$$

given:

$$\mathcal{X}^- = [\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_{\mathcal{N}_t - 1})]$$
$$\mathcal{X}^+ = [\mathbf{x}(t_2), \mathbf{x}(t_3), \dots, \mathbf{x}(t_{\mathcal{N}_t})].$$

The above operation is computationally quite expensive. However, when pre-processing the data through the SVD, DMD can be used to compute a low-rank operator $\widetilde{\mathcal{A}} \in \mathbb{R}^{r \times r}$, which is the projection of the true operator $\mathcal{A}$ onto the first r SVD modes of $\mathcal{X}$:

$$\widetilde{\mathcal{A}} = \widetilde{\mathcal{U}}^T \mathcal{A} \widetilde{\mathcal{U}} = \widetilde{\mathcal{U}}^T \mathcal{X} \widetilde{\mathcal{V}} \widetilde{\Sigma}^{-1}. \tag{6}$$

It has been mathematically proven [10] that the eigenvalues of $\widetilde{\mathcal{A}}$ are strongly connected to those of the true operator $\mathcal{A}$. Then, DMD can be used to estimate the dominant eigenvalues and eigenvectors associated to $\widetilde{\mathcal{A}}$, avoiding the (expensive) computation of the pseudo-inverse $\mathcal{X}^\dagger$. The procedure to find the operator $\widetilde{\mathcal{A}}$ is reported in Algorithm 1.

Input: Snapshot Matrix $\mathcal{X} = [\mathbf{x}(t_1) \mid \mathbf{x}(t_2) \mid \cdots \mid \mathbf{x}(t_{N_t})] \in \mathbb{R}^{\mathcal{N}_h \times N_t}$
Output: Best fit operators $\mathcal{A}$ and $\widetilde{\mathcal{A}}$

Generate time advancement matrices:
$\mathcal{X}^- = [\mathbf{x}(t_1) \mid \cdots \mid \mathbf{x}(t_{N_t - 1})] \in \mathbb{R}^{\mathcal{N}_h \times (N_t - 1)}$;
$\mathcal{X}^+ = [\mathbf{x}(t_2) \mid \cdots \mid \mathbf{x}(t_{N_t})] \in \mathbb{R}^{\mathcal{N}_h \times (N_t - 1)}$;

Compute the Singular Value Decomposition (SVD):
$\mathcal{X}^- = \widetilde{\mathcal{U}} \widetilde{\Sigma} \widetilde{\mathcal{V}}^T$;

Compute best fit operator: $\mathcal{A} = \mathcal{X}^+ \widetilde{\mathcal{V}} \widetilde{\Sigma}^{-1} \widetilde{\mathcal{U}}^* \in \mathbb{R}^{\mathcal{N}_h \times \mathcal{N}_h}$;

Project $\mathcal{A}$ onto the modes with rank r:
$\widetilde{\mathcal{A}} = \widetilde{\mathcal{U}}^* \mathcal{A} \widetilde{\mathcal{U}} \in \mathbb{R}^{r \times r}$

Algorithm 1: Dynamic Mode Decomposition: Computing the Best-Fit Linear Operator

Once the low-dimensional surrogate $\widetilde{\mathcal{A}}$ has been computed, the advancement in time is obtained from its spectral decomposition:

$$\widetilde{\mathcal{A}} \mathcal{W} = \mathcal{W} \Lambda, \tag{7}$$

where $\mathcal{W}$ and Λ are the eigenvectors and eigenvalues matrix respectively. From the former, it is possible to compute the high-dimensional DMD modes $\Phi \in \mathbb{R}^{\mathcal{N}_h \times r}$, that is, the eigenvectors of $\mathcal{A}$ [10]:

$$\Phi = \mathcal{X}^+ \widetilde{\mathcal{V}} \widetilde{\Sigma}^{-1} \mathcal{W} \tag{8}$$

Once the surrogate model has been defined, the solution can be reconstructed for unknown time values, following the expansion of Eq. (7):

$$\mathbf{x}_k \approx \sum_{j=1}^{r} \phi_j \lambda_j^{k-1} b_j, \tag{9}$$

with $\phi_j \in \mathbb{R}^{\mathcal{N}_h}$ the jth column of Φ, λ_j the jth diagonal element of Λ and b_j the mode amplitude, computed by projecting the initial condition onto the reduced space:

$$\mathbf{b} = \Phi^T \mathbf{x}_1 \tag{10}$$

2.2. Optimized DMD

One drawback of DMD is its difficulty in interpreting noisy datasets. Constant signals are often incorrectly interpreted as oscillatory, resulting in inaccurate spatio-temporal modes. This issue arises from the approximation of DMD through a linear system $\mathbf{x}' = \mathcal{A}\mathbf{x}$, whose exact solution is $\mathbf{x} = \mathbf{x}_0 \cdot e^{\mathcal{A}t}$. The evolution of the system strongly depends on the eigenvalue–eigenvector pairs $\omega - \Phi$ of $\mathcal{A}$, and the general solution can be written as $\mathbf{x} = \Phi \cdot e^{\omega t}\mathbf{x}_0$. Optimized DMD [11,12] can overcome this issue directly performing a regression to exponential-time dynamics by solving the following optimization problem:

$$\arg\min_{\omega,\Phi,\mathbf{b}} \left|\left| \mathcal{X} - \sum_{j=1}^{r} b_j \phi_j e^{\omega_j t} \right|\right| = \arg\min_{\omega,\Phi,\mathbf{b}} ||\mathcal{X} - \Phi e^{\omega t}\mathbf{b}||, \tag{11}$$

with r the dimension of the surrogate space in which the optimized solution is sought, Φ and ϕ_j the DMD modes, ω_j the continuous eigenvalues and $\mathbf{b}$ the amplitudes. Compared to standard DMD, which requires uniform sampling of the data, optimized DMD generalizes to non-uniform trajectories [12].

3. The magnetron injection gun

The Magnetron Injection Gun (MIG) is a key gyrotron component which generates a 2 MW electron beam through thermionic emission to produce a 1 MW radio-frequency wave (with 50% efficiency). Geometry details are in [4]: briefly, input power is provided by cladded filaments exchanging heat through radiation with the inner components, chiefly the emitter. Heat is focused by thermal shields delimiting the inner cavity. The structure is supported by the prolongator and nosecone, connecting to the heat-pipe where cooling oil flows. For the model description, the inner cavity includes filaments, cladding, emitter, and thermal shields; the outer cavity includes external components (prolongator, nosecone, connector, heat-pipe). Materials and properties are not reported for confidentiality (see Fig. 1).

The MIG operates in two distinct phases with different dynamics:

- On-phase: high voltages (~70 kV) are applied to the emitter, and additional power is provided to the filaments to compensate for the thermionic effect (Eq. (13)). The emitter and components heat up, and electron current is emitted into the resonant cavity.
- Off-phase: no voltage is applied and no current is produced. Filament power decreases to the steady-state value, cooling the inner components (while external components heat towards a new equilibrium).

The thermionic emission of the MIG is approximated with the Richardson–Dushman formula [13]:

$$J = \eta \cdot A_0 \cdot T^2 \exp\left(\frac{W_0 + c_1 T - a_s\sqrt{E}}{K_b T}\right), \tag{12}$$

with K_b the Boltzmann constant, T the emitter surface temperature in K, W_0 the emitter working function in eV, $c_1 T$ a temperature-dependent correction term for W_0 (in eV), E the local electric field (in V/cm), $a_s\sqrt{E} = \Delta U$ the correction on the working function due to the Schottky barrier lowering effect (in eV), with $a_s = 1.2$ eV·(cm/V)$^{1/2}$ the Schottky constant, and $A_0 = 120$ A/cm^2/K^2 is the (practical-units) Richardson constant, such that J is expressed in A/cm^2. The hyperparameter η, to be optimized, accounts for uncertainties related to the working function and the average heat extracted by each electron. Then, the heat extracted from the emitter surface can be computed as [14]:

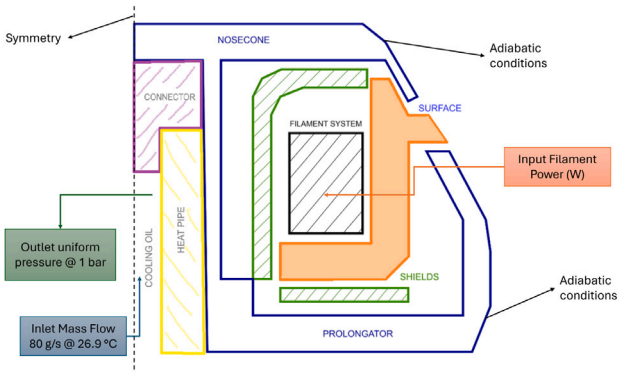

Fig. 1. Schematic representation of the MIG cathode (modified from [3]), along with the relevant boundary conditions. Upper and lower shields are shown in dashed green, while the emitter is marked in orange. The oil flows from the bottom of the central pipe upward to the connector n(purple, dashed), then it descends downwards in contact with the heat pipe (yellow, dashed). The MIG cathode is vacuum-sealed. The figure is not to scale.

$$q'' = (2K_b T + W_0 + c_1 T - \Delta U)J \tag{13}$$

This heat loss term leads to additional cooling of the system during the on-phase, if not balanced: during the on-phase additional power must be provided to the filaments.

4. MIG lumped model

The zero-dimensional lumped model describing the MIG thermal dynamics includes all components, from filaments to cooling oil, predicting thermal behaviour and beam current given accelerating voltage and input filament power. In this approximation, each body has a unique spatially uniform temperature, with temporal evolution described by heat equations coupled with energy balances. Thus, the heat exchange between components must be described in detail: the equivalent electric circuit describing heat transfer within the MIG is in Fig. 2.

In the inner cavity, the main heat transfer mode is radiation, along with conduction for components in contact. In the outer cavity, the main mode is conduction: additionally, convection occurs between the connector and cooling oil, and the heat-pipe and cooling oil. For each component i, the following energy balance holds [15]:

$$m_i c_i \frac{dT_i}{dt} = q_{in} - q_{out}, \tag{14}$$

with m being the mass, c the specific heat capacity, q_{in} the gained heat and q_{out} the lost heat. These terms depend on the heat transfer mechanisms between body i and all bodies it exchanges heat with. While some cases are straightforward, most heat exchange occurs within the inner and outer cavities, characterized by complex geometries with diffuse, grey, opaque surfaces separated by vacuum.

Moving outward from the filaments to the cooling oil, the most significant modelling features are listed below:

- The filaments exchange heat through radiation only with their cladding:

$$\dot{q} = \frac{\sigma(T_{a,i}^4 - T_{f,i}^4)}{\frac{1-\varepsilon_f}{\varepsilon_f A_f} + \frac{1}{A_f} + \frac{1-\varepsilon_a}{\varepsilon_a A_a}}, \tag{15}$$

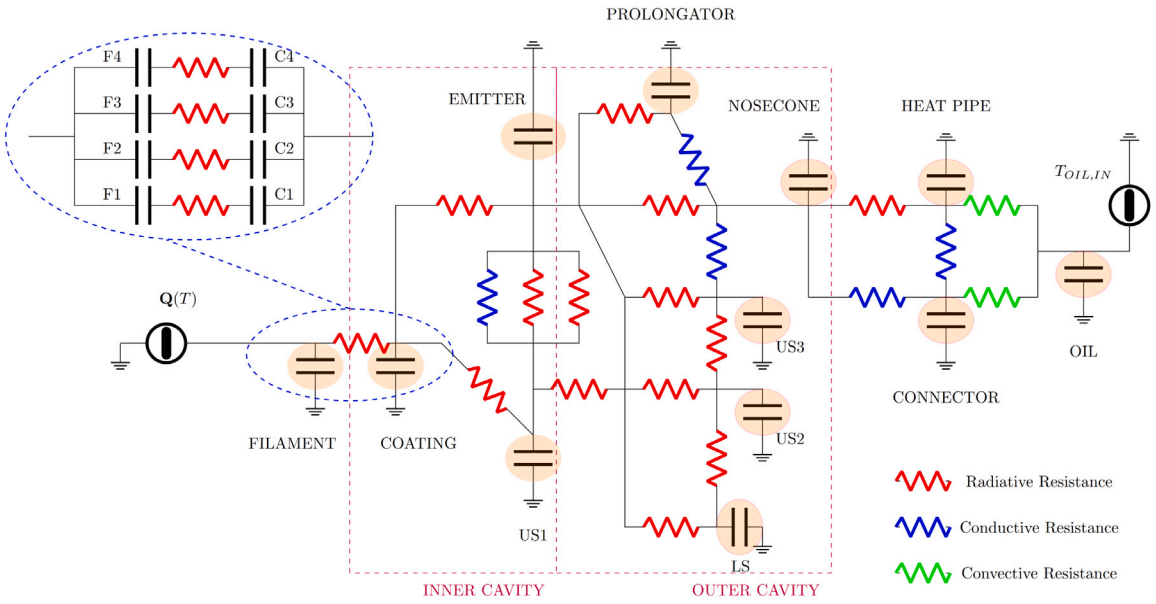

Fig. 2. Equivalent electrical circuit of the heat transfer between the various bodies in the MIG. Components are highlighted in orange. All components within a cavity exchange heat through radiation with all components within the same cavity. Resistances shown here are just for visualization purposes, to highlight the different heat transfer mechanisms between different components of the MIG.

with σ being the Stefan–Boltzmann constant, ε the body emissivity and A the surface area seen by the other body.

- The cladding dissipates heat in the inner cavity, and for each component in the inner cavity, the radiosity flux (the rate at which radiation leaves a surface per unit area) can be computed as:

$$J_i = R_i \cdot \sigma T_i^4$$
$$R_i = \left(\frac{1}{\varepsilon_i A_i} + 0.5 \sum_j A_i F_{i,j} - 0.5 A_i \right)^{-1} \frac{1}{\varepsilon_i A_i}$$
$$F_{i,j} = \frac{1}{A_i} \int \int_{A_i A_j} \frac{\cos\theta_i \cos\theta_j}{\pi R^2} d A_i d A_j \tag{16}$$

R_i is the radiosity matrix, converting surface black-body emission for body i into a net heat flux: the first term is radiative surface resistance $R_{s,i}$, whereas the second is geometric resistance. This term quantifies how shape and orientation affect radiative heat transfer from body i to j; $F_{i,j}$ is the view factor between bodies i and j, quantifying the fraction of surface j seen by surface i. For each body, the sum of view factors must be 1. From the radiosity flux, transferred heat is derived as:

$$\dot{q}_i = \frac{E_{b,i} - J_i}{\frac{1-\varepsilon_i}{\varepsilon_i A_i}}, \tag{17}$$

where $E_{b,i}$ is the black-body emission for body i.

- To capture the localized temperature variations on the emitter surface, this body is divided into three different regions: body, surface and intermediate connecting zone. These latter zones are described by four non-physical hyperparameters which must be optimized based on the available experimental data: two mass capacitances $C_{1,2}$ and $C_{2,3}$ for the thermal inertia and two thermal resistances $R_{1,2}$ and $R_{2,3}$ for the conductive heat transfer within the body. Conduction within the emitter is then modelled as:

$$\dot{q} = \frac{T_{e,i} - T_{e,i+1}}{R_{i,j}} \tag{18}$$

Once the emitter surface temperature has been computed, the beam current and the extracted heat can be computed using Eqs. (12) and (13).

- The heat transfer in the outer cavity is modelled following the same reasoning for the inner cavity. Heat transfer between components in contact is modelled following the same logic as Eq. (18): given the lack of information on surface roughness and contact pressure, the contact resistances (prolongator-nosecone $R_{p,n}$, nosecone-connector $R_{n,c}$ and connector-pipe $R_{c,h}$) are considered as hyperparameters to be optimized.
- The cooling oil exchanges heat with the connector and the heat pipe through convection. Based on the Reynolds number and oil flow characteristics obtained from a 3D conjugate heat transfer model of the MIG, two correlations were selected to evaluate the convective heat transfer coefficient h_i.
  - Pipe-oil: Graetz–Nusselt correlation for laminar flow in an annulus [16].
  - Pipe-connector: Sieder–Tate correlation for impinging flow [17]

  Once the convective heat transfer coefficient is known, the convective heat transfer can be computed as:

$$\dot{q} = h_i A_i (T_i - T_{oil}) \tag{19}$$

Given an experimental filament current and cathode voltage, the model predicts the cathode beam current and the temperature evolution of all included components, for a total of $\mathcal{N}_h$ 32 features, for both the on-phase and the off-phase; the number of time instants $\mathcal{N}_t$ depends on the time vector associated with the experimental input, thus it changes based on the considered pulse. For clarification, the typical model output is shown in Fig. 3 in terms of filament temperature (which can be considered fast, given the characteristic time constant of the order of 10^{-2} seconds and the dynamics immediately after the start of the pulse), emitter surface temperature and beam current: the strong difference between on-phase and off-phase dynamics can be clearly seen in the markedly-different temporal behaviour of the temperatures and especially in the current, which sharply drops to zero once the pulse is switched on. As an interesting note, the emitter temperature increases during the initial transient of the off-phase. This is because the main cooling effect given by the thermionic effect is missing from the emitter's energy balance during the off phase. This causes thermal

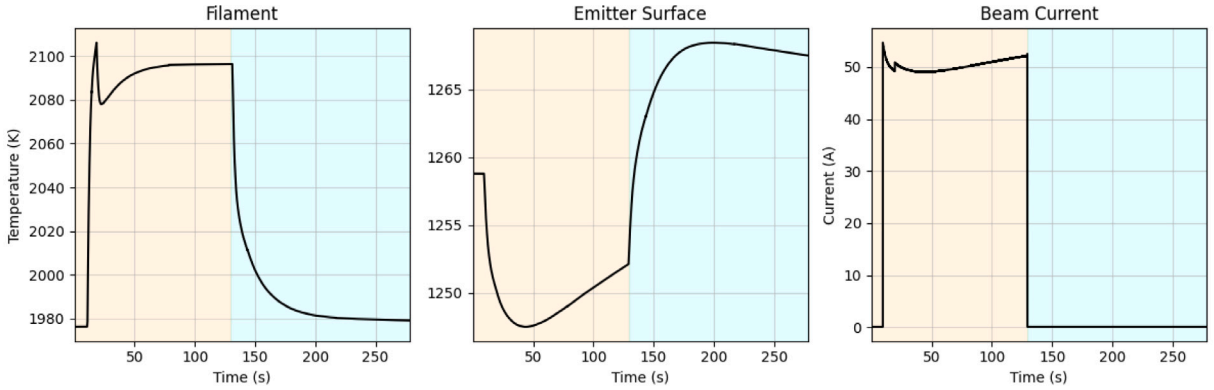

Fig. 3. Temporal behaviour of all quantities of interest of the MIG lumped model. The dataset includes both the on-phase (orange background, up to $t_1 = 129$ s) and the off-phase (cyan background). The different time scales of the different features, as well as the different dynamics of the two phases, can be clearly seen.

Table 1
Characteristic time constants and Biot numbers for the MIG components.

Component	τ (s)	Bi
Filament	0.0114	0.000636
Coating	0.0198	0.0791
Emitter	3.02	0.206
Shields	1.08 to 0.174	0.00001 to 0.001
Prolongator	1.05	0.329
Nosecone	8.65	0.272
Connector	24.3	0.0544
Heat pipe	56.3	4.32

energy from the surrounding components, which are warmer than the emitter, to diffuse to it. This results in an increase in the emitter's temperature during the first transient of the off-phase.

As can be seen in Fig. 3, the characteristics time constants range from 0.001 s for the filament and coating to tens of seconds for the external components (connector and heat-pipe). For completeness, the characteristic time constants, alongside the Biot numbers, are summarized in Table 1, where time constants are computed as follows:

$$\tau = \frac{m \cdot c_p}{h A_h}$$

with m the mass of the component, c_p the specific heat capacity, h the equivalent heat transfer coefficient and A_h the surface through which the body exchange heat.

To determine the optimal value of the unknown hyperparameters, an in-depth sensitivity analysis was carried out using the Differential Evolution (DE) algorithm; its details, as well as the final outcome, were reported in [18].

Following optimization, the model was validated against long pulses from FALCON. Summary results are in Fig. 4: lacking experimental temperature data, experimental beam current was compared with the model-computed values. Validation metrics include: (1) Normalized Root Mean Square Error (NRMSE), measuring average percentage error; (2) Mean Averaged Error (MAE), measuring average absolute point-wise error; (3) correlation coefficient (CC), measuring waveform similarity; (4) charge error (CE), measuring total delivered charge error; (5) Mean Absolute Percentage Error (MAPE), measuring mean relative error; (6) absolute difference at the last on-phase time-step (ϵ_f). The above quantities are computed as follows:

$$\text{NRMSE: } \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}}{\max(y) - \min(y)} \tag{20}$$

$$\text{MAE: } \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{21}$$

$$\text{CC: } \frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)\left(\hat{y}_i - \bar{\hat{y}}\right)}{\sqrt{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2 \sum_{i=1}^{N}\left(\hat{y}_i - \bar{\hat{y}}\right)^2}} \tag{22}$$

$$\text{CE: } \frac{\left|\int_0^T y(t)\,dt - \int_0^T \hat{y}(t)\,dt\right|}{\int_0^T y(t)\,dt} \tag{23}$$

$$\text{MAPE: } \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{24}$$

$$\epsilon_f = \left|y_N - \hat{y}_N\right|, \tag{25}$$

given the generic reference quantity y and its DMD estimator $\hat{y}$, N the total number of time steps, $\bar{y}$ the mean of the reference quantity, T is the total duration of the on-phase.

The model demonstrates good agreement across pulses with different input conditions (filament current and cathode voltage). Normalized RMSE remains below 15% for most pulses, with several below 10%, indicating high fidelity. Larger NRMSE values (pulses 9638, 9763) coincide with reduced correlation coefficients, indicating that these pulses exhibit both amplitude mismatches and wave form deviations. Conversely, pulses with larger NRMSE but high CC, such as pulse 9785, present errors that are primarily amplitude offsets, with the model correctly capturing the overall temporal trend, as confirmed in Fig. 5. The correlation coefficient (CC) exceeds 0.95 in most cases, capturing dominant temporal evolution. Regardless, the model performs well in both integral and point-wise metrics: charge error (CE) is typically below 3%, and peak error (PE) remains within 5%. Mean absolute percentage error (MAPE) stays below 5%, indicating small relative errors when current is non-negligible.[2] Overall, the model captures global beam properties (charge and peak current) and pulse timing, with moderate waveform deviations in specific cases. Fig. 5 reports the two worst estimated pulses: for pulse 9785, the shape is correct but the model underestimates current (final error ~1 A, maximum ~2 A); for pulse 9887, the shape is slightly distorted at later steps, overestimating final current by ~1 A. The reason why the current density decreases in the first transient in the figure is due to the thermo-ionic effect (see Eq. (13)). More heat is extracted through the emission of electrons than is absorbed by the emitter from the filaments. Consequently, the emitter temperature decreases during the initial transient, resulting in a reduction in the emitted current.

Finally, the lumped model requires roughly one minute to simulate both on-phase and off-phase for a given pulse; more detailed, higher-fidelity models may take between 5 h (2D) to nearly one day (3D

[2] Note that, whereas for all temperature features the metrics were computed across the whole temporal range, for current the metrics were computed only for the on-phase, being the current zero in the off-phase

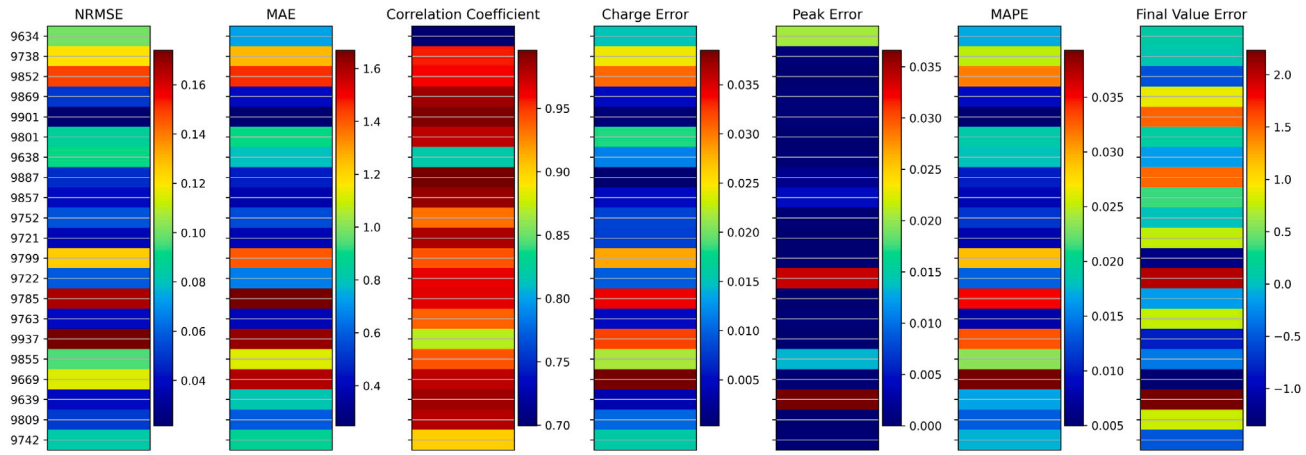

Fig. 4. Validation summary for the considered pulses (on-phase), using the following metrics: NRMSE, MAE, correlation coefficient, charge error, peak error, MAPE, absolute difference at the last time step.

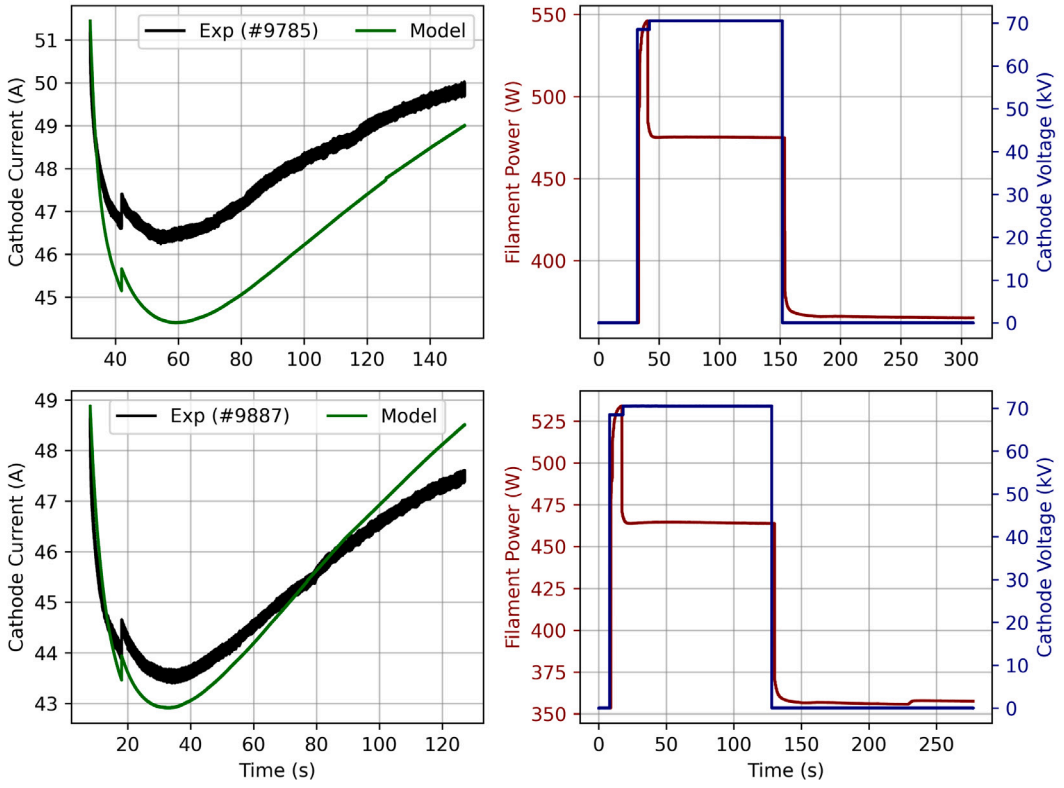

Fig. 5. The two worst estimated pulses in terms of overall error (left) and NRMSE (right): whereas the former indicates over/underestimation but correct shape, the latter indicates discrepancy in the current shape but overall good agreement in terms of current magnitude. Note that the experimental current reported here has been post-processed to reduce its noise, to better highlight its temporal behaviour.

with conjugate heat transfer) to simulate a single on-phase heating transient. This impressive time saving further justifies the choice of using a lumped model for this methodological work; regardless, the higher-fidelity models are currently being finalized and they will be used to create larger datasets to further assess the potentialities of Dynamic Mode Decomposition and, more in general, reduced order models for this application.

5. Data preprocessing and results

The dataset $\mathcal{X}$ includes 31 temperature fields (all MIG components) and the beam current ($\mathcal{N}_h = 32$); each field was sampled at $t = 0.001$ s, for $\mathcal{N}_t = 278,289$ snapshots. As seen in Fig. 3, features have different time scales: characteristic time constants span from less than seconds (innermost components) to tens of seconds (external components). Such

datasets are not optimal for DMD, which operates better on features with similar time scales. To recover a general surrogate model, $\mathcal{X}$ includes both the on-phase (voltage and beam power on, from $t_0 = 9$ s to $t_1 = 129$ s) and the off-phase (voltage and beam power off, from $t_1 = 129$ s to $t_2 = 278$ s). This implies the dataset includes two phenomena with different dynamics and temporal scales, adding complexity. Finally, uncorrelated Gaussian white noise is added to each feature separately; for a given feature $x(t)$, the noisy dataset $\mathcal{Y}$ contains:

$$y(t) = x(t) + \mathcal{N}(0, 1). \tag{26}$$

For this analysis, only fictitious noise is applied. However, in real applications within the digital twin framework, noise affects the data and it must be treated accordingly. A denoising algorithm based on Optimized DMD and Hankel embedding theory (Appendix A) is proposed in Appendix B.

When adopting machine learning and data-driven techniques, appropriate pre-processing significantly impacts algorithm performance. For computational speed, down-sampling is commonly used: following trials to find the largest ratio without information loss (omitted for brevity), a ratio of 10 was selected, leading to a dataset $\mathcal{X} \in \mathbb{R}^{\mathcal{N}_h \times \mathcal{N}_t/10}$. Finally, the dataset was scaled to level feature magnitudes: specifically, the `StandardScaler` from `scikit-learn` was used, rescaling features such that the scaled dataset $\mathcal{Z}$ has zero mean and unitary standard deviation:

$$z(t) = \frac{x(t) - \mu}{\sigma} \tag{27}$$

Clearly, this synthetic noise represents and idealization. In practice, sensors operating in the extreme electromagnetic environment of a megawatt-class gyrotron are likely subject to more complex noise sources, including high-frequency electromagnetic interference, non-stationary disturbances, and correlated noise components that deviate significantly from the zero-mean Gaussian assumption. However, the use of ideal white noise allows establishing a controlled test case for the denoising procedure and the DMD-based surrogate model in a well-defined, reproducible setting. The extension of the proposed framework to more realistic noise models will be a matter of future work.

Data denoising

Applying the procedure in Appendix B reduces the dataset's standard deviation by ~87% (from 1 to 0.13) while retaining the original dynamics of $\mathcal{X}$. Denoised features across different time scales are shown in Fig. 7. To quantify how well the procedure preserves dynamics, similarity metrics between the denoised dataset $\mathcal{Y}_{denoise}$ and clean reference $\mathcal{X}$ were computed. The mean squared error is very small (MSE = 0.0165) and the normalized MSE is negligible ($\mathcal{O}^{-9}$), indicating minimal energy distortion. Consistently, the global signal-to-noise ratio is high (SNR = 81.45 dB, up from 63.63 dB), confirming the noise residual is orders of magnitude weaker than the signal. To measure temporal pattern preservation, the Pearson correlation coefficient was computed: across all features, the mean is very high ($\bar{r} = 0.9703$), showing strong agreement with the original behaviour. The minimum value ($r_{min} = 0.1171$) corresponds to the oil temperature, which is constant in the original data; noise-induced oscillations around this constant reference explain the low correlation. All other features show $\bar{r} > 0.93$.

The denoised algorithm adopted is compared against other denoising techniques, including: Savitzky–Golay filter, low-pass Butterworth filter, Principal Component Analysis denoising, median filter, Gaussian smoothing. Selected metrics for the comparison are noise reduction, signal-to-noise ratio, average Pearson correlation coefficient and Mean Squared Error. Results are reported in Table 2.

Compared to Hankel-OptDMD denoising, the median filter provides a comparable noise reduction, showing that for this kind of noise, simple spike-removal strategies can be effective in preserving high-frequency features. Classical linear smoothing techniques, such as Savitzky–Golay, Butterworth and Gaussian filters record worse performance compared to Hankel-OptDMD, as they tend to slightly distort temporal dynamics. Conversely, PCA denoising reports the highest mean Pearson correlation coefficient (0.982), meaning that it preserves slightly better the temporal patterns; however, its noise reduction is lower, signifying that retaining a limited number of principal components reduced noise only partially. Overall, Hankel-OptDMD is capable of strong noise suppression whilst preserving the temporal fidelity: compared to simpler linear filters, its main drawbacks are its greater complexity from a mathematical standpoint, and slightly worse computational efficiency, especially as the Hankel matrix increases in size (see Fig. 6).

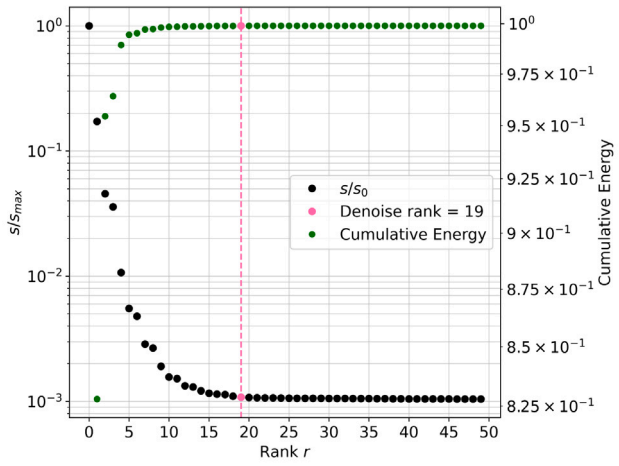

Fig. 6. Singular values of the embedded Hankel matrix (see Eq. (34)) (black) and cumulative energy (green). As a rule of thumb, the 'elbow' in the singular values separates physical from noisy features.

5.1. Optimum rank selection

When considering reduction problems for quasi-real-time applications such as digital twins, the trade-off between reconstruction accuracy (capability of the model on reconstructing the dynamics from a reduced space) and computational times (speed at creating and reconstructing data by the model) becomes significant. For DMD and its variants, the selection of the rank r of the surrogate linear space affects both: a higher-rank surrogate model can capture all the dynamics occurring in the system at the cost of a more expensive surrogate model generation time.

For the dataset $\mathcal{X}$, the optimal selection of the rank r can be formulated as a multi-objective optimization problem to minimize the $L2$ reconstruction error and the computational cost:

$$\epsilon_2(r) = \frac{||\mathcal{X} - \hat{\mathcal{X}}_r||_2}{||\mathcal{X}||_2} \tag{28}$$

$$t(r) = \text{wall time for fit + reconstruction} \tag{29}$$

with $\hat{\mathcal{X}}_r$ being the reconstructed dataset with OptDMD at rank r. As a preliminary indication, the Pareto front, the set of ranks where the error cannot be improved without increasing the time, and vice-versa, is shown in Fig. 8. Up to 20 s, the error drops by almost an order of magnitude, whereas beyond it drops only by a factor of 3, and a clear elbow can be seen around 30 s: beyond this point, the computational cost significantly increases for only small gains in reconstruction accuracy. Above the elbow, the first few points are scattered, an indication of under-resolved dynamics. The rank corresponding to the knee is $r = 6$, whereas the optimum value of the rank is $r = 12$.

Rank $r = 12$ corresponds to a total computational time of 41 s (of which the reconstruction time is less than 0.1 s) and a relative $L2$ error

Table 2
Metrics for the different denoising techniques. Desiderata are high noise reduction, mean Pearson coefficient and signal-to-noise ratio, and low Mean Squared Error.

Method	Noise reduction (%)	MSE	Mean Pearson r	SNR (dB)
Hankel-OptDMD	87.161	0.016	0.970	81.455
Savitzky–Golay	78.020	0.048	0.959	76.784
Butterworth	79.381	0.046	0.960	77.340
PCA denoising	59.088	0.167	0.982	71.388
Median filter	86.673	0.018	0.962	81.125
Gaussian filter	79.712	0.0412	0.959	77.480

of 7.61%. For prediction/forecast problems, the optimal choice of the rank also depends on the cutoff window (Section 5.4). Control-specific applications may also have additional requirements related to runtime and accuracy, depending on the safety-level. For example, the optimum rank to have reconstruction time less than 0.1 s and accuracy greater than 95% would shift to $r = 16$.

Fig. 9 reports the relative $L2$ error per feature considering optimum rank value $r = 12$. Reconstruction is significantly better for slow-scale dynamics, with error below 5% for all these features except for the two emitter surfaces; the error for the filaments (fast-scale dynamics) is around 10%, and the error for the beam current is roughly double. In fact, both filaments and beam current exhibit vertical jumps at the beginning and at the end of the on-phase. Since DMD-based algorithms assume linear Koopman evolution, they represent dynamics as smooth exponentials, meaning that the surrogate model is trying to approximate step changes using damped oscillatory modes, the best linear approximation of a jump-type discontinuity. By considering two different DMD models for the two different phases, ignoring the discontinuities, the errors for the filaments falls down to 2% (on-phase) and 4% (off-phase), whereas for the beam current in the on-phase the error is 8%. This is made cleared by Fig. 10, reporting the reconstructed filament and emitter temperature and beam current, for the case in which the full dataset is considered (red) and the case in which a dedicated surrogate model is built for the two phases separately (green), keeping $r = 12$: by separating the dynamics the jumps are hidden, and the two surrogate models predict a much smoother dynamics, closely resembling the actual field (black).

5.2. MIG latent dynamics identification

DMD can be used to extract the key temporal modes describing the system. As the dataset contains two different physical phenomena, external-driven growth and free decay, these are studied separately to capture the key physics of each regime, allowing for easier interpretation and comparison, with the only drawback being the loss of information regarding the transition. DMD modes represents the system' overall deviation from equilibrium, and not the local behaviour of specific components: as such, mode decay towards zero indicates that the system is approaching equilibrium; conversely, exponential growth indicates that the system is straying away from equilibrium.

The physical interpretation of each DMD mode refers to the known thermal behaviour of the MIG during each operating phase. Specifically, each mode is characterized by its eigenvalue, which encodes the associated damping time scale and oscillation frequency, and by its initial amplitude, which reflects the relative importance of that mode in the overall system response. By comparing these quantities against the known time constants of the MIG components and the expected thermal behaviour during the on-phase (externally driven heating) and off-phase (free cooling), each mode can be associated with a dominant physical process. This interpretation is clearly qualitative in nature, as the lumped model does not provide spatial resolution: nonetheless, it is consistent with the governing physics and the observed temporal dynamics in Fig. 3.

On-phase. The normalized (with respect to the initial condition) temporal modes for the on-phase are reported in Fig. 11, along with their initial amplitudes. The dominant mode (blue) is non-oscillatory, and physically it represents the global temperature field produced by the input power: this is the bulk thermal response of the system, and the monotonic behaviour indicates heat addition. The large initial amplitude refer to a forced response of the system, causing the large initial jump in the filaments temperature; the transient part is an exponentially decaying deviation from the final equilibrium, indicating that eventually the system will reach a final equilibrium state. The second mode (red), the quasi-steady thermal mode, represents the temperature increase of the large external components, whose time constant is long compared to the duration of the on-phase: physically, it captures the slow accumulation of heat leading to an asymptotic steady-state. Modes three (green) and four (purple) are the thermal diffusion eigenmodes, which capture intermediate-time responses (all internal components bar the filaments and the radiation heat exchange between inner components). The remaining modes capture the step-like response of the filaments to the input, as well as the small initial oscillations in the temperature profiles of some components.

Off-phase. The normalized dynamics modes of the off-phase are reported in Fig. 12. Modes one (blue) and four (purple) decay very slowly, and they represents the heat stored in the inner cavity, which gets redistributed to the large external components, causing them to heat up during the off-phase. These two modes dominate the long-term asymptotic behaviour, and they indicates that the MIG will slowly reach the new equilibrium state. Modes two (red), three (green) and six (black) are the intermediate heat diffusion modes, which capture mid-range thermal processes such as the redistribution through the thermal shields, the radiative exchange between internal surfaces, and the emitter surfaces reaching a uniform equilibrium temperature, equal to that of the emitter body. This is indicated by the smooth exponential decay and the moderate damping time. Finally, modes five (orange), seven (pink) and eight (light blue) represents the fast redistribution of heat between the filaments and the cladding, which at the start of the off-phase rapidly adjust their temperature towards the new equilibrium state. Note that in thermal systems, oscillations represents complex-conjugate pairs associated with rapid transient reshaping of the temperature field.

5.3. MIG spectral characteristics

Since DMD computes a low-dimensional surrogate of the dynamical matrix $\mathbf{A}$, having the same eigenvalues, any consideration derived from the analysis of the eigenvalues of this surrogate $\widetilde{\mathbf{A}}$ can be applied straightforwardly to the true system matrix. The eigenvalues of $\widetilde{\mathbf{A}}$ are reported in Fig. 13, for both the on-phase and the off-phase. Each eigenvalue $\omega = \mathcal{R}(\omega) + i\mathcal{I}(\omega)$ encodes a damping/amplification time scale (real part) and a oscillation frequency (imaginary part). Only a few modes exhibit mild oscillatory behaviour: as oscillatory pairs always appear as complex-conjugates, they do not correspond to physical resonances, but they represent rapid reshaping of the temperature field between components with different time constants. In fact, their oscillatory periods ($2\pi/|\mathcal{I}(\omega)| \approx$ 4–30 s) coincide with the observed short-term 'thermal adjustments' following the step change (filament to cladding heat diffusion, early redistribution within the inner cavity

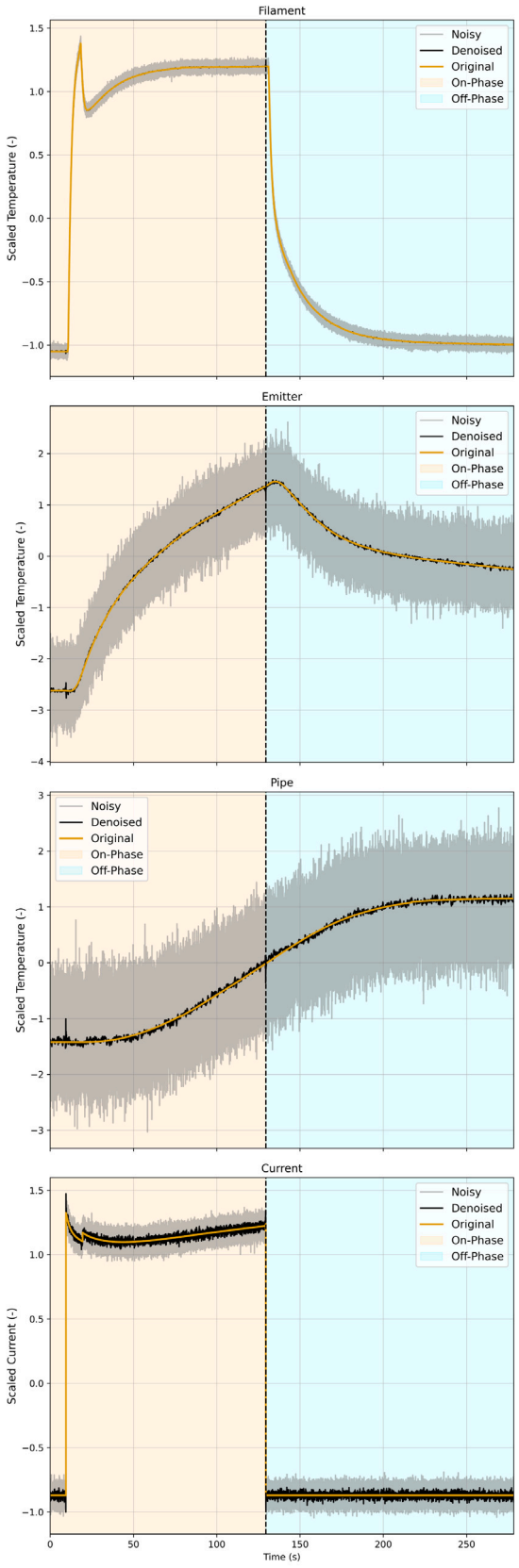

Fig. 7. Comparison between the noisy data (black), the denoised data following the reported procedure (gold) and the original, clean data (dashed line) for three different time dynamics.

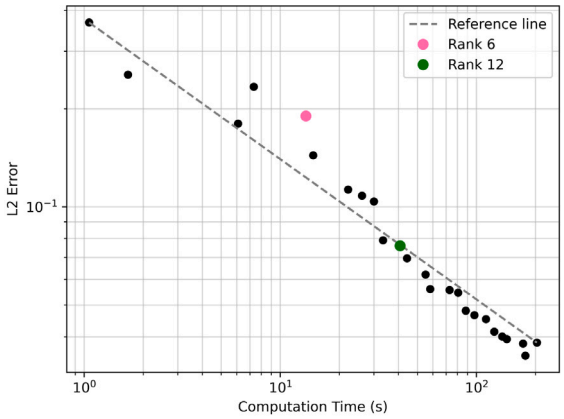

Fig. 8. Pareto front of the multi-objective optimization problem of the selection of the optimal rank, comparing computational times (x-axis) and overall relative $L2$ error (y-axis). Both knee value (pink) and optimum rank (green) are reported as well, with respect to the reference line.

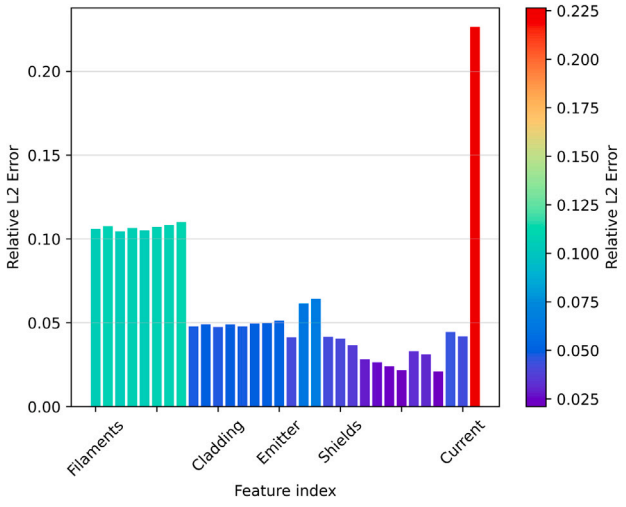

Fig. 9. Relative $L2$ error per feature considering optimum rank $r = 12$.

and the emitter body). Considering the real part, strongly damped modes with $\mathcal{R}(\omega) \approx -0.2 : -0.4$ represents fast transients, which dominates the initial seconds of both the on-phase and the off-phase. Modes clustered around $-0.02 : -0.05$ represent slow thermal processes with characteristic times of tens of seconds. Finally, modes with $\mathcal{R}(\omega) \approx 0$ correspond to the large external components. In fact, during the on-phase the whole system slowly approaches a new equilibrium temperature, whereas during the off-phase the same slow mode becomes the dominant cooling component.

The slightly positive on-phase eigenvalue $\omega_1 \approx 2.4 \times 10^{-4}$ s^{-1}, it does not indicate numerical instability or physical divergence, as observed in Section 5.2. Instead, it indicates that the on-phase represents the slow, externally-driven heating, with energy being injected continuously and the global temperature of the system monotonically increasing towards the new equilibrium state: once the input is removed, the mode moves to the negative real axis, as expected from passive cooling.

5.4. MIG dynamics forecast

Another relevant task in the digital twin framework is the capability to extrapolate the state of the system in time beyond the available

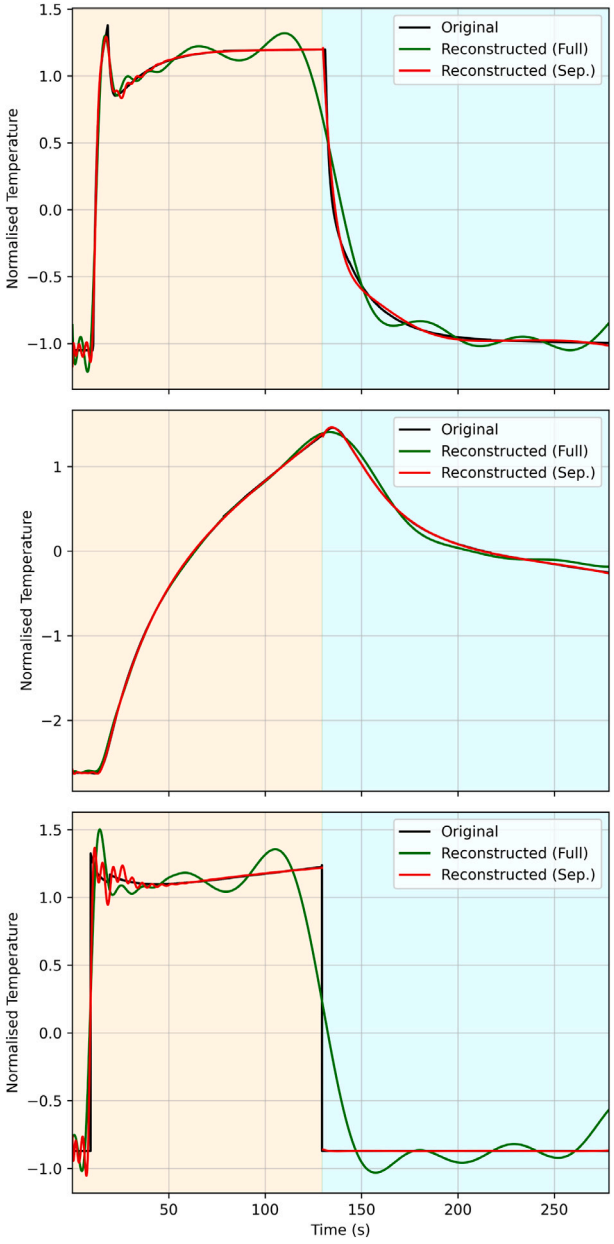

Fig. 10. Feature reconstruction for the filament, emitter and beam current for the case in which a single surrogate model accounts for the entire time range (green) and the case in which on-phase and off-phase are separated, and two independent surrogate models are built. All models have rank $r = 12$.

training range (i.e. forecasting). The focus is on predicting with acceptable accuracy the final state of the system over the time window of interest, to perform informed decisions for control. Given their different dynamics, the two phases are studied separately, building a dedicate surrogate model for each: for sake of brevity, and given its relevancy for control, only the results for the on-phase are reported here. The multi-objective optimization problem for the selection of the optimal rank now must also consider the cut-off training threshold, that is, the final time instant of the dataset used for training the model. Therefore, as a first step, the optimization of the dimension of the surrogate model is performed, prioritizing lowering as much as possible the cut-off training time. It is worth mentioning that the results obtained in this section refer to the single pulse the model has been trained upon: generalization to multiple pulses, which falls into the domain of parametric reduced order methods, is the primary direction for future works.

The optimization analysis has been carried out as follows: first, only ranks with overall $L2$ reconstruction error lower than 5% have been considered in the analysis ($r \in [7 : 31]$); for each of these rank values, the cut-off training threshold has been lowered to find the minimum value satisfying some constraints. Additionally, the slope β of the experimental and model signal has been computed, to ensure that the trend is captured correctly. The constraint on the final prediction error for feature y_i is then[3]:

$$|\sigma_i - \hat{\sigma}_i| \leq 2 \quad \text{and} \quad |\beta_i - \hat{\beta}_i| \leq 0.03 \tag{30}$$

with σ_i being the standard deviation associated with feature i and β_i the slope of the signal: only if this constraint is satisfied for all the features, the prediction is considered acceptable. As such, the validity of the forecast is assessed through these two quantifiable criteria, which measure respectively the absolute deviation between the forecast and the true signal at the last time step and the accuracy of the predicted trend. A forecast is considered valid if both criteria are simultaneously satisfied for all features of interest.

Fig. 14 shows the minimum allowed cutoff training value for different values of the rank. Generally, as the rank increases, the minimum required training fraction decreases until rank $r = 24$, which corresponds to cutoff of 79%. Beyond this point, there is no substantial reduction in the required training fraction: instead, higher ranks increase model capacity but also sensitivity, meaning that more training data are needed to reliably keep all features below the error threshold; for example, high-rank models may fit noise in the training data.

For the case with $r = 24$ and training cutoff 79%, the maximum forecast error at the last time step is 1.86 °C, for the cladding, whereas the average error is 1.01 °C; for the beam current, the error is 0.71 A. For all different ranks the filaments and the cladding are the components exhibiting the largest errors, likely due to their fast dynamics, especially at the beginning of the transient: in fact, immediately after the jump at the start of the on-phase, the filaments exhibit a sharp decrease of around 20 °C; this dynamics gets encoded in the DMD model, which learns that after an increase there may be a decrease in temperature (in fact, the forecast always underestimates the true temperature). The reconstructed and forecast temperature for the worst-forecast filament, worst-forecast cladding and the beam current are reported in Fig. 15. Not only the constraint (30) is satisfied, but for all cases, the forecast remains within the oscillations due to noise.

The more relaxed the constraints are, the more the cutoff training value may shift towards lower values. For example, considering that up to now there are no available on-phase experimental data on temperature, and that the most important output of the model is the beam current, the constraint may be applied only on this feature, for example requiring a forecast error on the beam current at the last time step lower than the noise affecting the observations and that the overall trend is captured with reasonable accuracy. The 'soft' constraint now becomes[4]:

$$-4.4\ \text{A} \leq |\sigma_i - \hat{\sigma}_i| \leq 4.4\ \text{A} \quad \text{and} \quad |\beta_i - \hat{\beta}_i| \leq 0.03 \tag{31}$$

Fig. 16 shows the optimum results with this 'soft' constraint, obtained with rank $r = 9$ and with a minimum train cutoff of 0.62. Higher ranks lead to worse performance, likely indicating over-fitting, and to properly capture the dynamics more data are needed. The model with

[3] Given the lack of experimental data on temperature and of a consensus of the required accuracy for digital twins for fusion components, constraints for temperature were chosen based on the authors' expertise.

[4] For current, given the availability of experimental data, constraints were chosen so that the forecast is considered acceptable if it remains within the experimental uncertainty bands.

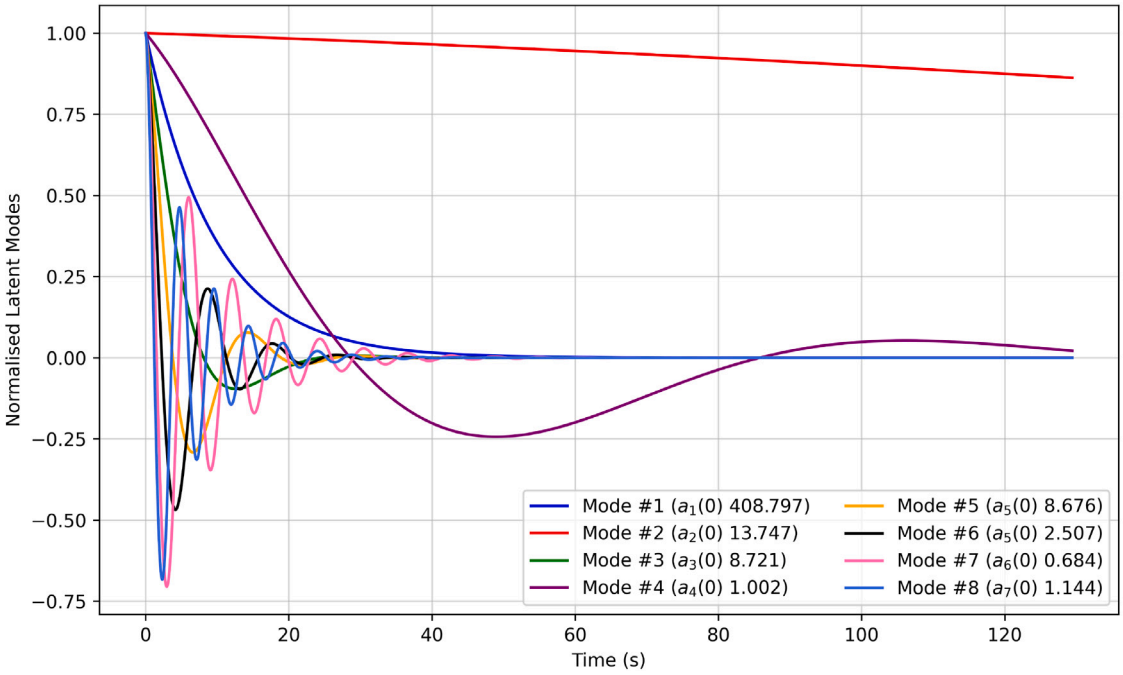

Fig. 11. First eight temporal modes for the on-phase, ranked according to the magnitude of the associated singular value.

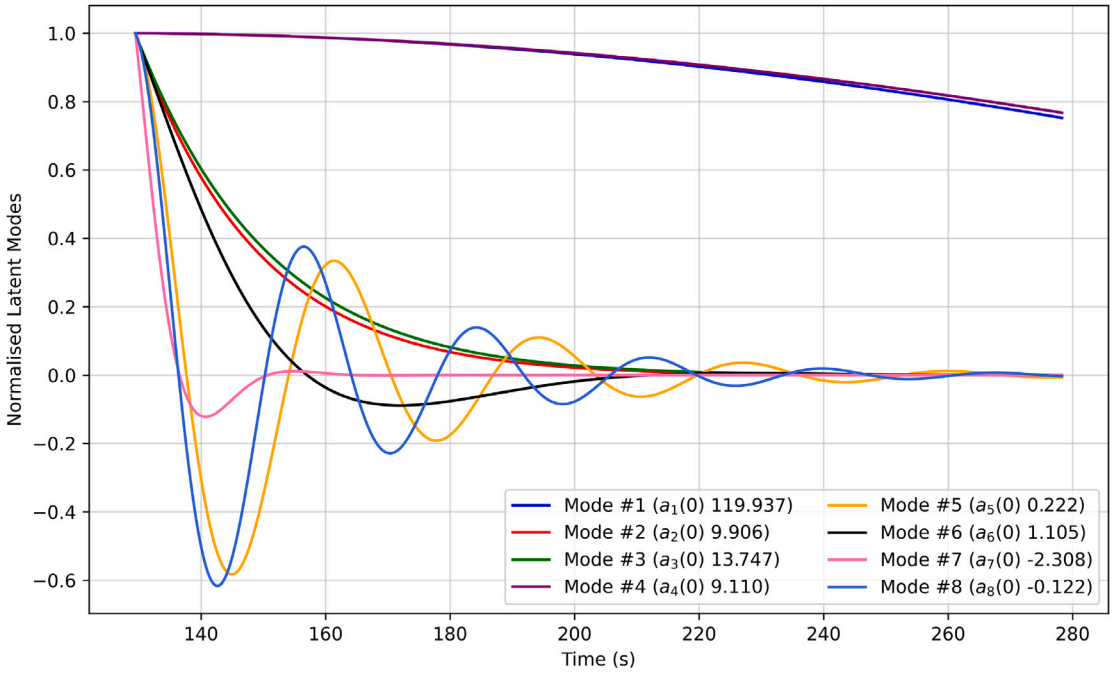

Fig. 12. First eight temporal modes for the off-phase, ranked according to the magnitude of the associated singular value.

$r = 9$ suffers in the reconstruction phase, approximating the vertical jumps with sinusoidal signals, however the performance in forecast mode is quite good, with a final error of 1.36 A and a slope error of 2.26%. This model overall runs in 23 s (including both training and forecast phase: the latter takes less than 0.1 s).

6. Conclusions

In this work, Dynamic Mode Decomposition (DMD) was successfully applied to identify the key transient dynamics of the Magnetron Injection Gun (MIG), establishing a robust methodological framework for digital twin applications in fusion energy systems. The challenge of balancing computational efficiency with predictive accuracy is critical for digital twin applications, especially for a component like the MIG which is characterized by complex non-linear dynamics and vastly different temporal scales, ranging for less than seconds for filament response to tens of seconds for external components dynamics. Given the methodological nature of the work, the starting model is a validated lumped model of the MIG, to better focus the analysis on DMD characteristics and performance.

The results show that DMD, combined with Hankel embedding, provides an effective data pre-processing strategy, achieving significant

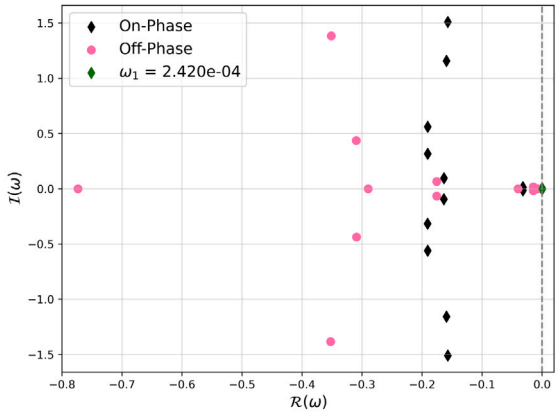

Fig. 13. Eigenvalues of the surrogate matrix $\tilde{\mathbf{A}}$, considering separately the on-phase (black) and the off-phase (pink). The magnitude of the least stable mode is reported in green.

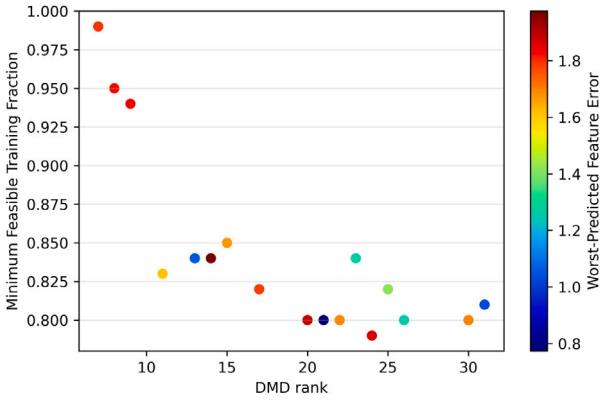

Fig. 14. Minimum allowed cutoff training value for different ranks, under constraint (30) for all features of interest. The maximum absolute error is reported as well.

noise reduction while preserving the essential temporal features of the system across multiple time scales. This step proved crucial for separating fast physical dynamics from uncorrelated noise and for improving the robustness of the reduced-order representation. Once trained, the DMD-based surrogate model successfully extracted the dominant temporal modes governing the MIG thermal behaviour during both the on-phase and the off-phase, offering a clear physical interpretation of the underlying processes. These include the global thermal response driven by input power, the slow heating of large external components, intermediate diffusion mechanisms, and fast filament dynamics, thus revealing the hierarchical structure of heat transfer within the device.

The spectral analysis of the reduced-order operator further demonstrated that the DMD surrogate retains meaningful stability information. The eigenvalues were found to be predominantly non-oscillatory, with the slightly positive eigenvalue observed during the on-phase correctly reflecting the externally driven heating process, while the off-phase dynamics were characterized by stable, purely dissipative behaviour. This spectral insight is particularly relevant for monitoring and control-oriented applications.

In reconstruction mode, an optimal rank of 12 provided a good compromise between accuracy and computational cost, yielding a relative $L2$ error of 7.61% with total execution times compatible with quasi-real-time operation. In addition, the surrogate model showed promising forecasting capabilities. When strict accuracy requirements were imposed on all state variables, a higher-rank model trained on

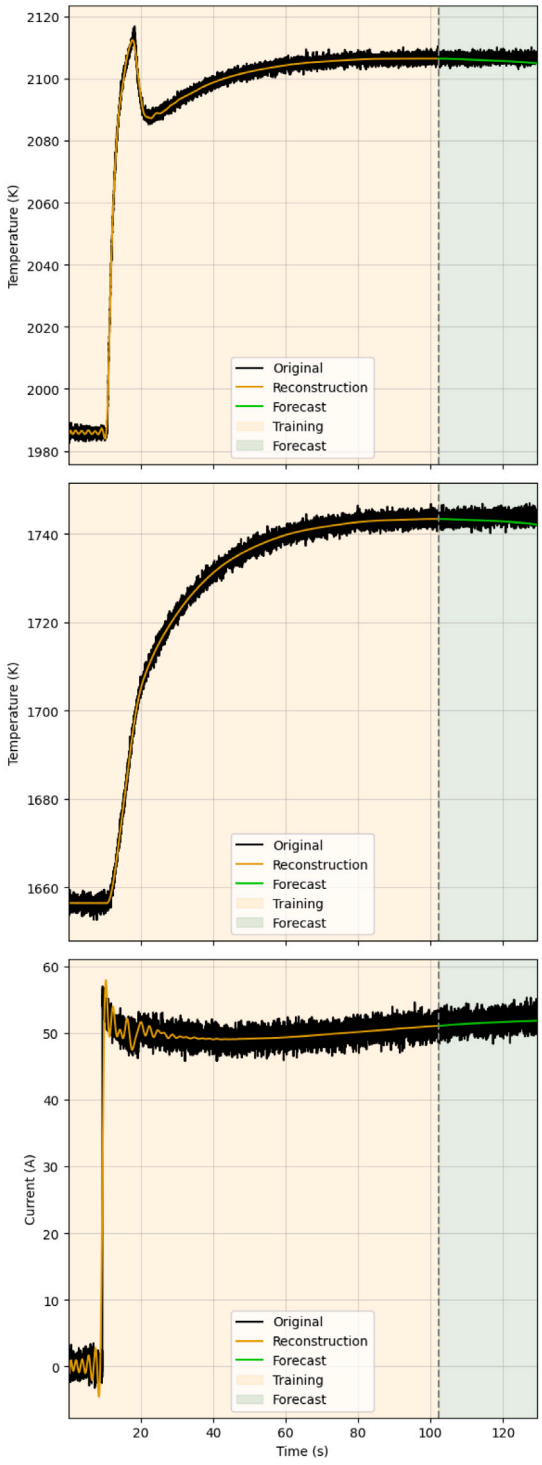

Fig. 15. Reconstructed (gold) and forecast (green) features (from top to bottom: filament, cladding, beam current) for the case $r = 24$ and cutoff training value 79%.

approximately 80% of the data accurately predicted the remaining dynamics. When the focus was restricted to the beam current, which is the only experimentally available measurement, reliable forecasts were achieved with lower-rank models and shorter training windows, further reducing computational requirements.

Regarding the forecast analysis, it must be mentioned that the DMD model was trained and validated on a single pulse, and its

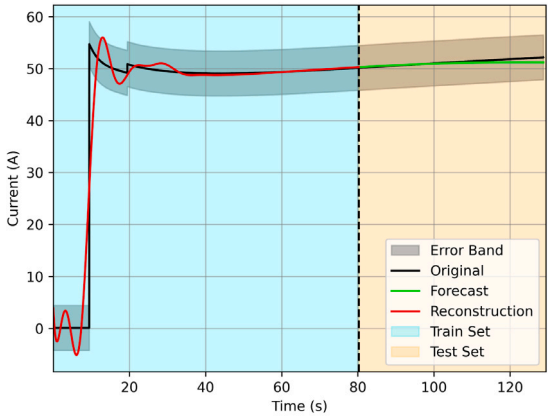

Fig. 16. Reconstruction and forecast of the beam current for rank $r = 9$ and train cutoff 0.62.

generalization to different operating conditions has not yet been assessed. Additionally, the cutoff training value was optimized purely on the basis of reconstruction accuracy, without direct correlation to the physical response time of the gyrotron system. Establishing such a correlation would require experimental data on system response times under operational conditions, which is currently unavailable.

Thus, this paper demonstrates how DMD can be used to build accurate and fast-computing surrogate models of fusion components, establishing a strong foundation for its application to more complex fusion component modelling. Future works will include: the extension of the analysis to parametric dataset, examining if the surrogate model is capable of correctly predicting unseen transients; the use of high-fidelity models for training data generation to identify the key spatial characteristics of the MIG; advanced stability and control analysis starting from the surrogate operator eigenvalues, using for example non-modal stability analysis; coupling of DMD with uncertainty quantification techniques. The idea is then to conduct further research into the reduced order modelling technique for developing the gyrotron digital twin, which is an ongoing project and more information are provided in [2].

CRediT authorship contribution statement

Elia Novarese: Writing – review & editing, Investigation. **Carolina Introini:** Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation. **Laura Savoldi:** Writing – review & editing, Supervision. **Antonio Cammi:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Hankel embedding

In linear algebra, a Hankel matrix is a structured matrix having constant skew diagonals, meaning that element $a_{i,j}$ depends only on $i + j$:

$$\mathcal{H} = \begin{bmatrix} a_1 & a_2 & \dots & a_m \\ a_2 & a_3 & \dots & a_{m+1} \\ . & . & . & . \\ . & . & . & . \\ a_d & a_{d+1} & \dots & a_{m+d-1} \end{bmatrix} \tag{32}$$

For DMD, these matrices are constructed by stacking time-shifted copies of a time series $x(t)$, so that each column of $\mathcal{H}$ is an m-length time-delay embedding of the signal. Such matrices are widely used in system identification [19] and singular spectral analysis [20] and are closely related to delay embedding techniques. Delay embedding encodes the state of a dynamical system through the time history of its observables, relying on the fact that the state of an n-dimensional system can be reconstructed from n independent observables [21,22]. In this embedded space, deterministic dynamics manifest as low-rank structures in the Hankel matrix, while noise remains high-rank. Consequently, the singular value decomposition of the Hankel matrix enables separation of coherent dynamics from noise, and reconstruction from the dominant modes effectively de-noises the signal [23].

Appendix B. Data denoising

Experimental data are usually affected by sensor noise, which makes it necessary to distinguish between high-frequency phenomena caused by noise and physical phenomena.

The steps of the proposed denoising algorithm are reported below:

1. To remove static biases and centre the data, the temporal mean of each feature i is subtracted from the dataset:

$$\mu_i = \frac{1}{\mathcal{N}_t} \sum_{t=1}^{\mathcal{N}_t} \mathcal{Y}(:,t) \rightarrow \tilde{\mathcal{Y}} = \mathcal{Y} - \mu \tag{33}$$

2. The centred data are embedded into a block-Hankel matrix (see A) $\mathcal{H} \in \mathbb{R}^{(\mathcal{N}_h \cdot w) \times (\mathcal{N}_t - w + 1)}$ with window size w.[5] In practice, the Hankel matrix is built for each window shift $i = 0, \dots, w-1$:

$$\mathcal{H}[i \cdot \mathcal{N}_h : (i+1)\mathcal{N}_h, :] = \tilde{\mathcal{Y}}(:, i : i + k - 1)$$
$$k = \mathcal{N}_t - w + 1 \tag{34}$$

3. The Hankel matrix is decomposed through Singular Value Decomposition to determine its low-rank structure, and the denoising rank r_n is chosen based on the singular values (1). The relevant modes are selected based on cumulative energy considerations, which is defined as:

$$\frac{\sum_{i=1}^{k} s_i^2}{\sum_{i=1}^{r} s_j^2}. \tag{35}$$

For denoising purposes (i.e., removing noise-dominated high-order modes while preserving all dynamically relevant features) the selected rank r_n is the one able to preserve almost 99.9% of the variance of the cumulative energy.

4. The Hankel matrix is now projected onto the dominant singular subspace, spanned by the first r_n modes:

$$\mathcal{H}_{proj} = \Sigma_r \mathcal{V}_r^T \in \mathbb{R}^{n_r \times \mathcal{N}_t} \tag{36}$$

5. Optimized DMD is applied to learn a dynamical model $\hat{H}(t)$ in the reduced space by predicting the denoised reduced-rank trajectory $\hat{H}_{proj}$

$$\hat{\mathcal{H}} = \mathcal{U}_r \hat{H}_{proj} \tag{37}$$

6. The denoised Hankel matrix is mapped back to the original time series dataset via averaging over overlapping segments, and the feature means are added back:

$$\tilde{\mathcal{Y}}_{denoised}(:, i : i + k - 1) + =$$
$$+ = \hat{\mathcal{H}}[i \cdot \mathcal{N}_h : (i+1)\mathcal{N}_h, :],$$
$$\text{for each } i = 0, \dots, w-1, \tag{38}$$

followed by point-wise normalization by the Hankel window and the addition to μ.

[5] For this analysis, $w = 40$.

Data availability

The data that has been used is confidential.

References

[1] M.I. Battye, S. Perinpanayagam, Digital twins in fusion energy research: Current state and future directions, IEEE Access 13 (2025) 75787–75821.

[2] E. Novarese, A. Cammi, R. Difonzo, C. Introini, L. Savoldi, An integrated simulator towards a digital twin for MW-class gyrotrons for fusion reactors, Fusion Eng. Des. 218 (2025) 115157.

[3] A. Antonione, R. Bertazzoni, G. Carannante, K. Cindric, M. Ferrari, A. Leggieri, F. Sanchez, G. De Tommasi, Development and benchmarking of a thermal lumped model for the magnetron injection gun of the MW-power ITER European gyrotron, Nucl. Fusion 65 (2025) 036034.

[4] N. Badodi, A. Cammi, A. Leggieri, F. Sanchez, L. Savoldi, A new lumped approach for the simulation of the magnetron injection gun for MegaWatt-class EU gyrotrons, Energies 14 (2021) 2068.

[5] G. Rozza, M. Hess, G. Stabile, M. Tezzele, F. Ballarin, C. Grassle, M. Hinze, S. Volkwein, F. Chinesta, P. Ladeveze, Y. Maday, A. Patera, J. Farhat Char, Model Order Reduction: Volume 2: Snapshot-Based Methods and Algorithms, De Gruyter, 2020.

[6] S.L. Brunton, J.N. Kutz, Data-Driven Science and Engineering: Machine Learning, Dynamical Systems and Control, Cambridge University Press, 2022.

[7] J.N. Kutz, S.L. Brunton, B.W. Brunton, J.L. Proctor, Chapter 3: Koopman analysis, in: Dynamic Mode Decomposition, SIAM, 2016, pp. 39–53, http://dx.doi.org/10.1137/1.9781611974508.ch3, URL https://epubs.siam.org/doi/abs/10.1137/1.9781611974508.ch3. _eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611974508.ch3.

[8] P.J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28.

[9] F. Faraji, M. Reza, A. Knoll, J.N. Kutz, Data-driven local operator finding for reduced-order modeling of plasma systems, J. Phys. D: Appl. Phys. 58 (2025) 175201.

[10] J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, J.N. Kutz, On dynamic mode decomposition: Theory and applications, J. Comput. Dyn. 1 (2014) 391–421.

[11] T. Askham, J.N. Kutz, Variable projection methods for an optimized dynamic mode decomposition, SIAM J. Appl. Dyn. Syst. 17 (2018) 308.

[12] D. Sashidhar, J.N. Kutz, Agging, optimized dynamic mode decomposition for robust, stable forecasting with spatial and temporal uncertainty quantification, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 308 (2022) 20210199.

[13] F.M. Charbonnier, R.W. Strayer, W. Swanson, E.E. Martin, Nottingham effect in field and T-F emission: Heating and cooling domains, and inversion temperature, Phys. Rev. Lett. 13 (1964) 397–401.

[14] W.B. Nottingham, Remarks on energy losses attending thermionic emission of electrons from metals, Phys. Rev. 59 (1941) 906–907.

[15] F. Incropera, T.L. Bergman, D.P. DeWitt, A.S. Lavine, Fundamentals of Heat and Mass Transfer, John Wiley & Sons, 2011.

[16] K. Hirbodi, M. Yaghoubi, D.M. Warsinger, New Nusselt number correlations for developing and fully developed laminar flows in concentric circular annular ducts, Int. Commun. Heat Mass Transfer 134 (2022) 105936.

[17] C. Sleicher, M. Rouse, A convenient correlation for heat transfer to constant and variable property fluids in turbulent pipe flow, Int. J. Heat Mass Transfer 18 (5) (1975) 677–683.

[18] C. Introini, N. Badodi, R. Bertazzoni, G. Carannante, J.P. Hogge, S. Illy, A. Leggieri, F. Sanchez, L. Savoldi, A. Cammi, Sensitivity analysis of the lumped thermal model of the EU 170 GHz gyrotron magnetron injection gun, IEEE Trans. Plasma Sci. 52 (2024) 3836–3841.

[19] M. Fazei, T.K. Pong, D. Sun, P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, SIAM J. Matrix Anal. Appl. 34 (2013) 946–977.

[20] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, Phys. D. 35 (1989) 395–424.

[21] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Geometry from a time series, Phys. Rev. Lett. 45 (1980) 712–716.

[22] F. Takens, Detecting strange attractors in turbulence, in: Dynamical Systems and Turbulence, 1981, pp. 366–381.

[23] P. Frame, A. Towne, Space-time POD and the Hankel matrix, PLoS One 18 (2023) e0289637.