



Data-driven model order reduction for sensor positioning and indirect reconstruction with noisy data: Application to a Circulating Fuel Reactor

Antonio Cammi^{*}, Stefano Riva, Carolina Introini, Lorenzo Loi, Enrico Padovani

Politecnico di Milano, Department of Energy, CeSNEF-Nuclear Engineering Division, Nuclear Reactors Group, via La Masa, 34 20156 Milano, Italy

ARTICLE INFO

Dataset link: ERMETE Lab, <https://github.com/ERMETE-Lab/ROSE-pyforce>

Keywords:

Hybrid Data-Assimilation
Generalized Empirical Interpolation Method
Indirect Reconstruction
Sensors positioning
Molten Salt Fast Reactor
Noisy data

ABSTRACT

Sensor positioning and real-time estimation of non-observable fields is an open question in the nuclear sector, especially for advanced nuclear reactors. In Circulating Fuel Reactors (CFR), liquid fuel and coolant are homogeneously mixed, and thus these reactors will not have internal structures, making sensor positioning in the primary circuit, including the core, an unresolved problem, making most of the core blind to sensors. Thus, the possibility of estimating the system state in the whole domain using a few local measurements has important implications for safety, monitoring, and control both in nominal and accidental conditions. In this context, the integrated Model Order Reduction and Data Assimilation framework offers intriguing opportunities to reliably combine experimental data and background knowledge from a reduced mathematical model. This work discusses and applies innovative methods within this framework, based on the Generalized Empirical Interpolation and the Indirect Reconstruction algorithms, to a proposed concept of CFR. This work aims to identify the optimal sensor positioning within the core and assess the feasibility of reconstructing the quantities of interest starting only from transient sparse data on fuel temperature, possibly noisy, and testing the predictive capabilities of the discussed methods.

1. Introduction

The problem of accurate state estimation is a key issue from the point of view of safety, control, and monitoring (Argaud et al., 2018; Introini, 2021). Typically, two main sources of information are available: a mathematical model of the system and experimental observations. The former is limited by modelling assumptions and by the values of the parameters characterizing the system, affected by uncertainty; the latter is polluted by random and systematic noise, and it is sparse in terms of coverage of the whole domain of interest (Introini et al., 2023b).

Data Assimilation (DA) (Carrassi et al., 2018; Brunton and Kutz, 2022) techniques allow for the integration of the two possible sources of information. Traditional DA methods are based on optimization problems: they require to have multiple evaluations of the mathematical model to find the optimal prediction. However, accurately solving several times parametrized Partial Differential Equations (PDEs) usually involves numerical methods, which may have very high computational costs. Model Order Reduction (MOR) (Lassila et al., 2014; Quarteroni et al., 2015; Rozza et al., 2020) techniques aim at reducing the burden of multiple numerical simulations by replacing the high-fidelity problem with one featuring a lower complexity, to perform faster simulations whilst ensuring the conservation of the characteristic features of the physical phenomena under consideration. Amongst MOR

techniques, the non-intrusive ones offer a handy framework to couple mathematical models with experimental observation, and accordingly, a hybrid approach between MOR and DA can be pursued. Literature works on this field, also applied to complex physical and engineering systems, are now briefly reported: Daescu and Navon (2007) introduced a general framework for MOR and DA applied to meteorological problems, Maday et al. (2015) proposed a Generalised Empirical Interpolation Method for the Stokes equation, also considering inverse modelling (thus the reconstruction of the velocity from pressure measurements), then Introini et al. (2023b), Riva et al. (2023a) and Riva et al. (2024) applied the previous techniques to fluid dynamics and nuclear reactors test cases; the works of Cao et al. (2021) and Pei et al. (2020), which have adopted data-driven modelling by Neural Network, can be included in this similar framework. The approach proposed in this paper focuses on linear MOR techniques, which are suited for the type of problem under study and are more physically consistent compared to Neural Network regression.

Some MOR techniques offer a framework in which to address both the problem of optimal sensor positioning and real-time reconstruction of the state of the system, as they can work with incomplete datasets to simulate the sparsity of experimental sensors (Gong et al., 2021).

^{*} Corresponding author.

E-mail address: antonio.cammi@polimi.it (A. Cammi).

Some other MOR techniques start from the knowledge of some solutions of a parametrized model to build an empirical interpolant by selecting the interpolation points (i.e., ‘sensors’) that best approximate the given high-fidelity solutions (Maday, 2006; Gong et al., 2016). As mentioned above, experimental observations are polluted by random and systematic noise. On the other hand, simulation data can be considered affected by noise as well: in this case, noise may refer to model uncertainties, numerical errors, discretization errors and so on, which affect the accuracy of the simulation result compared to the *true* solution (Maday et al., 2014; Riva et al., 2024). Many MOR techniques do not have algorithms to deal with noise, as they assume that the experimental observation and/or the high-fidelity solution snapshot is the ‘truth’; not only that, but when data are affected by noise some MOR algorithms lose their good properties, and the approximation error diverges (Argaud et al., 2017) as the noise is amplified to the point that adding more sensors or increasing the size of the reduced space worsen the reconstruction of the quantities of interest. Thus, stabilization techniques are needed.

Within these hybrid approaches, this work focuses on the Generalized Empirical Interpolation method (GEIM) (Maday et al., 2015; Maday and Mula, 2013; Maday et al., 2016) and the Indirect Reconstruction (IR) (Introini et al., 2023b). GEIM works by approximating the state of the system through the simultaneous combination of measurements coming from external sensors and a reduced parametrized model. Considering the issue of noisy measurements and/or snapshots, a novel implementation of the GEIM algorithm developed by the authors in Introini et al. (2023a), based on Tikhonov regularization and therefore called TR-GEIM, is used in this work. For what concerns IR, the algorithm developed by the authors in Introini et al. (2023b) is used.

Issues with this method arise when only partial observations are available, i.e., one or more fields of interest cannot be directly detected because the corresponding set of physical sensors is not available. Rigorously speaking, the absence of data makes data assimilation not possible, however for strongly coupled problems such as Circulating Fuel Reactors (CFRs) investigating the possibility of extracting some information on the non-observable fields from the observable ones is legitimate (Introini et al., 2023b; Gong et al., 2022). Not only in CFRs temperature and velocity are interdependent, but also the precursors’ density and the neutron flux strongly depend, respectively, on velocity and temperature, and these fields will contain some information about the others (Aufiero et al., 2014). This work will explore the possibility of reconstructing the whole state of a proposed CFR concept using only temperature sensors, both located in the whole core domain and along the boundaries only. The methods described in this work are general, and they can be applied regardless of the available set of sensors and the observable quantities.

This work starts from the work done in Maday et al. (2015) and expands the application done in Introini et al. (2023b), where the GEIM was applied to the Molten Salt Fast Reactor (MSFR) to reconstruct the neutron flux starting from temperature data following an accidental transient. Two different algorithms will be used: the performance of the two approaches mentioned before (GEIM and IR) will be compared. Compared to Introini et al. (2023b), this work aims at reconstructing the velocity and the precursors’ concentration starting from temperature data, and at predicting the transient evolution of the system and the new steady-state reached after stabilization, starting from a limited amount of data in time.

The paper is organized as follows: starting from a general overview of MOR methodology, Section 2 will be devoted to the description of the GEIM and IR methods; Section 3 will then briefly describe the considered case study; in the end, Section 4 will present the key results of the study and in Section 5 the main conclusions are drawn.

2. Methodology

Most of the MOR techniques are characterized by an offline-online decomposition (Lassila et al., 2014; Rozza et al., 2020; Brunton and Kutz, 2022) (Fig. 1). During the former, the MOR method solves several times the high-fidelity model to build a training set of snapshots (with \mathcal{N}_h degrees of freedom), which are used to build the reduced space onto which the model is projected, whose dimension $N \ll \mathcal{N}_h$. In this step, the positions of the sensors are also determined. This offline step is computationally expensive, but it is performed only once. In the online step, the computationally cheap reduced model built in the previous one, which can be a ‘small’ system of Ordinary Differential Equations (ODEs) or even a linear system, is solved for all required instances; in addition, measurements can be integrated to enhance the information of the mathematical background (Maday et al., 2014).

All methods that deal with indirect reconstruction and parameter estimation share the common hypothesis of bijectivity of the mapping between the parameter space and the observable quantity and vice versa (Introini et al., 2023b):

$$\mu \longrightarrow T(\mu) \quad \text{and} \quad T(\mu) \longrightarrow \mu. \quad (1)$$

This assumption cannot be easily proved a priori and hence in most of the cases it is simply assumed; according to numerical experience (Introini et al., 2023b; Riva et al., 2023b,a) this hypothesis is generally valid, even though there may be issues when bifurcation problems, strong non-linearities, hysteresis or steady state conditions are encountered. This implies that knowing the parameter $\mu \in D$ and the set of observations $T(\mu)$ allows for the reconstruction of the whole system state. For this work, temperature data serve as input for the data-driven parameters estimation problem, whose solutions serve as input for the MOR algorithm to reconstruct the unobservable fields.

As mentioned in the previous section, the selected MOR approach allows the selection of the interpolation points, which corresponds to the locations in the domain where the built map assumes the same value as the observation. From the physical point of view, these locations are where the sensors should be located to better approximate the whole domain of interest given a limited number of observations.

2.1. Generalized empirical interpolation method

The Generalized Empirical Interpolation Method (Maday and Mula, 2013; Maday et al., 2015) builds hierarchical spaces to approximate the given function with the interpolant:

$$u(\mathbf{x}; \mu) \simeq \mathcal{I}_M[u](\mathbf{x}; \mu) = \sum_{m=1}^M \beta_m(\mu) \cdot q_m(\mathbf{x}), \quad (2)$$

where the spatial behaviour is captured by the magic functions $\{q_m\}_{m=1}^M$ and the parametric dependence by the coefficients $\{\beta_m\}_{m=1}^M$, and where M is the dimension of the reduced space and μ is the parameters vector. The reduced basis functions are generated through a greedy procedure (Maday et al., 2015), which requires the solution manifold \mathcal{U} of the functions u , the parameter space D , and a library Y of functionals $v : \mathcal{U} \longrightarrow \mathbb{R}$, which represents the physical sensors (as they have physical dimensions and therefore cannot be approximated by a single value: rather, they are represented by local averages based on their centre of mass and their point spread). When the position of the sensors is not fixed (as in sensor positioning optimization problems), their location (the magic points) is selected through a greedy selection procedure.

GEIM uses an iterative greedy procedure which at every time step builds a reduced space. The offline procedure takes as inputs the available snapshots (i.e., the solution of the high-fidelity problem or some processed experimental data) and returns the set of magic functions and magic points $\{v_m\}_{m=1}^M$ by minimizing at every time step the interpolation error (Fig. 2). Thus, the magic sensors define the best

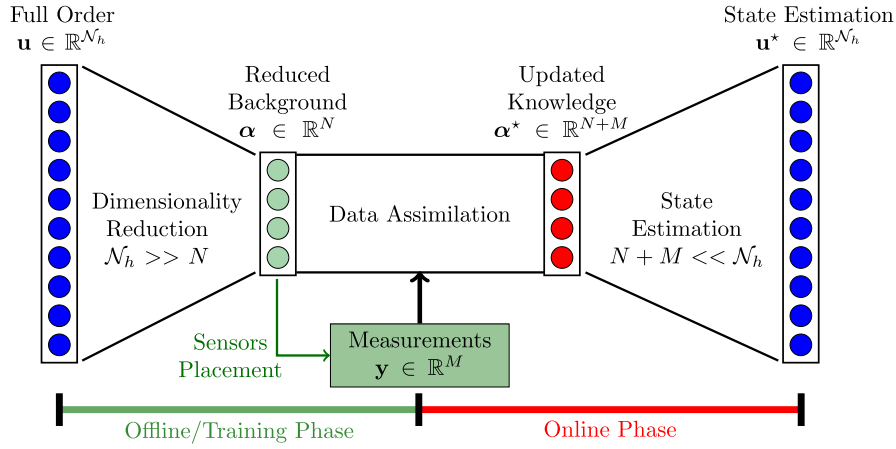


Fig. 1. General idea behind Hybrid Data-Assimilation techniques.

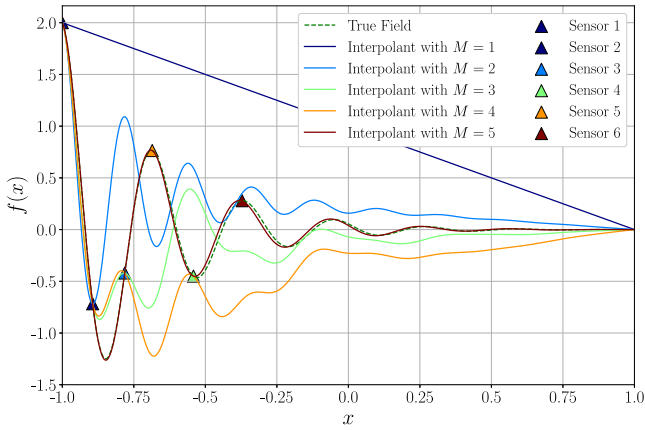


Fig. 2. Rationale behind the positioning of the sensors: at each time step, the algorithm selects the location corresponding to the largest interpolation error among all the available ones.

positions to maximize the amount of information extracted by the experimental setup. The overall computational procedure is summarized in Algorithm 1.

During the online phase, the data are collected:

$$y_m = v_m(u^{\text{true}}(\mathbf{x}); s) + \epsilon_m, \quad \text{with } m = 1, \dots, M, \quad (3)$$

where ϵ_m represents the disturbance on the data and s represents the spread of the sensor (a measure of the resolution of the physical sensor). This term has a strong impact on the interpolation error: when considering noise-free measurements, the interpolant is unique and the interpolation error is bounded (Maday et al., 2016); however, introducing noise on the data makes the interpolation error diverge (Argaud et al., 2017), as the GEIM algorithm in its standard formulation is not asymptotically robust against noise. This work uses a stabilized version of the GEIM, based on Tikhonov regularization, developed by the authors in Introini et al. (2023a) and labelled TR-GEIM, which, during the online phase, solves the linear system:

$$(\mathbb{B}^T \mathbb{B} + \lambda \mathbb{T}) \beta = \mathbb{B}^T \mathbf{y} + \lambda \mathbb{T} \langle \beta \rangle, \quad (4)$$

where $\mathbb{B}_{ij} = v_i(q_j)$ is the matrix of GEIM which contains the coefficients $\{\beta_m\}_{m=1}^M$, \mathbb{T} and λ are, respectively, the regulation matrix and the regularization parameter and $\langle \beta \rangle$ is the sample mean of the training coefficients. This regularization method is based on the inclusion of *a priori* information about the solution within the regulation matrix: this

Algorithm 1: GEIM greedy procedure - Offline Stage

Input

Maximum number of iterations M_{\max} ;

Tolerance δ ;

Training Snapshots $\{u^{(i)} = u(\mathbf{x}, \mu_i)\}_{i=1}^{N_s}$ with $\mu_i \in \Xi_{\text{train}}$;

Output

Magic functions $\{q_1(\mathbf{x}), \dots, q_M(\mathbf{x})\}$;

Magic sensors $\{v_1, \dots, v_M\}$;

Initialization

$M = 1$, $E_1 = \delta + 1$;

First iteration

$$u_1 = \underset{i=1, \dots, N_s}{\operatorname{argmax}} \|u^{(i)}\|_{L^2(\Omega)};$$

$$v_1 = \underset{v^{(k)} \in Y}{\operatorname{argmax}} |v^{(k)}(u_1(\mathbf{x}); \mathbf{x}_k, s)|;$$

$$q_1(\mathbf{x}) = \frac{u_1(\mathbf{x})}{v_1(u_1(\mathbf{x}))};$$

while $(M < M_{\max} \ \& \ \delta \geq E_{M-1})$ **do**

$M = M + 1$;

$$u_M(\mathbf{x}) = \underset{i=1, \dots, N_s}{\operatorname{argmax}} \|u^{(i)} - \mathcal{I}_{M-1}[u^{(i)}]\|_{L^2(\Omega)};$$

$$\text{given } \mathcal{I}_{M-1}[u^{(i)}] = \sum_{m=1}^{M-1} \beta_m(\mu) \cdot q_m(\mathbf{x});$$

$$\text{and } \sum_{m=1}^{M-1} \beta_m \cdot v_{m'}(q_m) = v_{m'}(u^{(i)}) \text{ with } m' = 1, \dots, M-1;$$

$$v_M = \underset{v^{(k)} \in Y}{\operatorname{argmax}} |v^{(k)}(u_M - \mathcal{I}_{M-1}[u_M]; \mathbf{x}_k, s)|;$$

$$q_M(\mathbf{x}) = \frac{u_M(\mathbf{x}) - \mathcal{I}_{M-1}[u_M](\mathbf{x})}{v_M(u_M - \mathcal{I}_{M-1}[u_M])};$$

$$E_M = \underset{i=1, \dots, N_s}{\operatorname{argmax}} \|u^{(i)} - \mathcal{I}_M[u^{(i)}]\|_{L^2(\Omega)};$$

matrix must be chosen properly, and can be written as

$$\mathbb{T}_{ij} = \begin{cases} \sigma_{\beta_i}^{-2} & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (5)$$

in which the penalisation applied to each coefficient depends on its sensitivity to noisy data, the greater the dispersion of $\{\beta_m\}_{m=1}^M$, measured by the variance $\sigma_{\beta_i}^2$ of training coefficients, the lower the impact of noisy data on it. The regularization parameter λ is an hyperparameter of the model and must be tuned according to the measurement noise. The complete procedure for the online phase is reported in Algorithm 2: more information about the mathematical background of the GEIM algorithm, its regularization using Tikhonov and the error theory behind the GEIM approach can be found in Introini et al. (2023a).

Algorithm 2: (TR-)GEIM - Online Stage**Input**

Magic functions $\{q_1(\mathbf{x}), \dots, q_M(\mathbf{x})\}$;
 Magic sensors $\{v_1, \dots, v_M\}$;

Output

State Estimation $\mathcal{I}_M[u^{\text{true}}](\mathbf{x})$;

Online Estimation

Acquisition of Experimental Data

$$\{y_m = v_m(u^{\text{true}}) + \epsilon_m\}_{m=1}^M;$$

Solve for β from the regularized problem (4);

Estimate the state of the system

$$u^{\text{true}} \simeq \mathcal{I}_M[u^{\text{true}}](\mathbf{x}) = \sum_{m=1}^M \beta_m \cdot q_m(\mathbf{x})$$

In the GEIM, an *a priori* error estimate of the interpolation procedure is given by the Lebesgue constant (Maday et al., 2016), defined as follows

$$\Lambda_M = \sup_{\phi \in \mathcal{U}} \frac{\|\mathcal{I}_M[\phi]\|_{\mathcal{U}}}{\|\phi\|_{\mathcal{U}}}. \quad (6)$$

The Lebesgue constant can be thought as a measure of the deviation from optimality of the hierarchical sub-spaces generated through the greedy procedure (Maday et al., 2016): then, the interpolation error scales linearly with Λ_M . Simply speaking, the higher the Lebesgue constant is, the further the generated subspace is from optimality, however, as there exist an upper bound for this quantity, the reduced order error is *a priori* bounded, regardless of the sub-optimality of the generated space. This quantity increases as more sensors are added, thus providing an upper bound also on the number of allowed sensors.

2.2. Indirect reconstruction

Being $\{q_m(\mathbf{x})\}_{m=1}^M$ the i -th the magic function generated by the GEIM for the observable field u , in the hypothesis of a one-to-one correspondence between parameters and the set of coefficients $\{\beta_m(\mu)\}_{m=1}^M$, the parametric state of the system can be retrieved from the coefficients. For any new parameter μ not belonging to the training set, the expansion coefficients are interpolated, and the mapping \mathcal{F} between parameters and coefficient is built. Then, if this map is correct, the parametric state can be estimated by looking for the value of μ such that $\mathcal{F}(\mu) = \beta(\mu)$. To guarantee that $\beta(\mu)$ belongs to the image of $\mathcal{F}(\mu)$, which is a numerical approximation of the real one, a minimization problem must be introduced. The idea is to minimize the discrepancy between the field reconstructed by interpolating the reduced coefficients and the experimental observations:

$$J^*(\mu) = \sum_{m=1}^M [v_m(\mathcal{I}_M[u](\mathbf{x}; \mu)) - y_m]^2 = \|\mathbb{B}\beta(\mu) - \mathbf{y}\|_2^2. \quad (7)$$

Then, the minimization problem becomes:

$$\hat{\mu} = \underset{\mu \in \mathcal{D}}{\operatorname{argmin}} J^*(\mu) = \underset{\mu \in \mathcal{D}}{\operatorname{argmin}} \|\mathbb{B}\beta(\mu) - \mathbf{y}\|_2^2, \quad (8)$$

such that $\beta \leftarrow \mathcal{F}(\mu)$. The constraints are now placed on the reduced coefficients. To solve this optimization problem, global minimization methods are needed. Other than GEIM, this work will also test IR using the Proper Orthogonal Decomposition with Interpolation (POD-I) (Demo et al., 2019), which uses the POD method in the offline phase to extract the reduced basis and the coefficients for the non-observable field ϕ . Being $\{\psi_m\}_{m=1}^M$ the set of POD basis, the high-fidelity solution $\phi(\mathbf{x}; \mu)$ is expressed as

$$\phi(\mathbf{x}; \mu) \simeq \sum_{m=1}^M \alpha_m(\mu) \cdot \psi_m(\mathbf{x}), \quad (9)$$

where $\{\alpha_m\}_{m=1}^M$ are the POD modal coefficients, given by the projection of $\phi(\mathbf{x}; \mu)$ onto each POD mode. The matrix of coefficient thus become, for each parameter in the training set

$$A_{pm} = (\phi(\mathbf{x}; \mu_p), \psi_m)_{\mathcal{U}}, \quad \mu_p \in \Xi_{\text{train}} \quad \text{and} \quad m = 1, \dots, M, \quad (10)$$

with $(\cdot, \cdot)_{\mathcal{U}}$ being the inner product in the space \mathcal{U} and $\Xi_{\text{train}} \subset \mathcal{D}$ being the training set of parameters. Then, in the online phase the modal coefficients are interpolated for any new parameter, obtaining an approximated solution. Different interpolation methods can be used: this choice severely impacts the interpolation error. It is worth mentioning that the developed IR algorithm based on GEIM is robust against noise, as it considers the minimization between the reconstructed field and the experimental observations. The complete procedure of the method adopted in this work is given in Algorithm 3: more details about the GEIM-based IR method can be found in Introini et al. (2023b) and Riva et al. (2023b).

Algorithm 3: Indirect Reconstruction Computational Procedure**Input**

Snapshots $(u, \phi)(\mathbf{x}; \mu) \quad \mu \in \Xi_{\text{train}}$;

Max Sensors Available M , Reduced space tolerance tol ;

Output

State estimation of the non-observable field $\phi^*(\mathbf{x}; \mu^*)$;

Offline Stage

Construction of the reduced spaces

$$Z_M^u = \operatorname{span}(q_1, q_2, \dots, q_M) = \text{GEIM-Offline}(u(\mu), \Xi_{\text{train}}, M, tol);$$

$$Z_N^\phi = \operatorname{span}(\psi_1, \psi_2, \dots, \psi_M) = \text{POD-Offline}(\phi(\mu), \Xi_{\text{train}}, M, tol);$$

Selection of sensors

$$v_m(\cdot) = \text{GEIM}(Z_M^u) \quad m = 1, \dots, M;$$

Construction of the maps $\mathcal{F} : \mu \rightarrow (\beta^u, \alpha^\phi)(\mu)$;

Online Stage

Collect measurements \mathbf{y} of field u ;

Estimation of the parameter μ^* by solving (8);

Compute the coefficients from the map $\mathcal{F} : \mu^* \rightarrow \alpha^\phi(\mu^*)$;

Estimation of the non-observable fields ϕ^* with (9);

3. Numerical case: the MSFR

As an example of a CFR, the Th-cycle based MSFR is used (Serp et al., 2014). The MSFR concept can be very useful in assessing the capabilities of integrated MOR with DA, due to the strong multi-physics coupling between neutronics and thermal hydraulics. This CFR uses a mixture of ^7LiF and ThF_4 (77.5 – 22.5 mol%), with a melting temperature of 838 K. In nominal operating conditions, the salt enters from the bottom at a temperature of 923 K and is extracted from the top of the core at a maximum temperature of 1023 K, with a nominal average in-core temperature of around 973 K. The total fuel salt volume in the core is 9 m³. Radially, the core is surrounded by a fertile mixture to improve breeding, not considered in this work. Since the goal is to show the capabilities of the methods discussed in Section 2 for sensor positioning and indirect reconstruction, for sake of simplicity the reference geometry used in this work is a 2D axisymmetric wedge (of amplitude 5°) of the EVOL MSFR geometry (Serp et al., 2014). The high-fidelity model is developed and solved using an OpenFOAM multi-physics model for transient analysis (Aufiero et al., 2014). This model uses the one-speed diffusion approximation for neutronics, convective-diffusive delayed and decay heat precursors, and Reynolds-Averaged Navier–Stokes equations (not reported here for sake of brevity) with

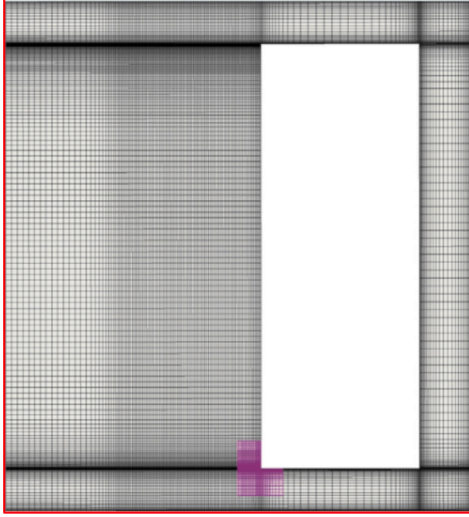


Fig. 3. Spatial mesh adopted (the purple region represents the core inlet, whereas the red lines depict the external boundary) (Aufiero et al., 2014).

energy for thermal-hydraulics:

$$\begin{cases} \frac{1}{v} \frac{\partial \phi}{\partial t} = \nabla \cdot (D \nabla \phi) - \Sigma_a \phi + (1 - \beta_t) v \Sigma_f \phi + \sum_{i=1}^8 \lambda_i c_i \\ \frac{\partial c_i}{\partial t} = -\mathbf{u} \cdot \nabla c_i + \nabla \cdot \left(\frac{v_i}{S_{c_i}} \nabla c_i \right) - \lambda_i c_i + \beta_i \cdot v \Sigma_f \phi \\ i = 1, \dots, 8 \\ \frac{\partial d_j}{\partial t} = -\mathbf{u} \cdot \nabla d_j + \nabla \cdot \left(\frac{v_j}{S_{c_j}} \nabla d_j \right) - \lambda_{h,j} d_j + \beta_{h,j} \cdot E_f \Sigma_f \phi \\ j = 1, \dots, 3 \\ \frac{\partial T}{\partial t} = -\mathbf{u} \cdot \nabla T + \nabla \cdot \left(\frac{K_s + K_t}{\rho C_p} \nabla T \right) - (1 - \beta_{h,t}) \frac{E_f}{\rho C_p} \Sigma_f \phi + \sum_{j=1}^3 \frac{\lambda_{h,j}}{\rho C_p} d_j \end{cases} \quad (11)$$

Both cross-sections (for reaction y) and the diffusion coefficient depend on temperature:

$$\begin{cases} \Sigma_y(T, p) = \frac{\rho}{\rho_0} \cdot \left[\Sigma_{y,0} + \alpha_{\Sigma_y} \log \left(\frac{T}{T_0} \right) \right] \\ D(T, p) = \frac{\rho}{\rho_0} \cdot \left[D_0 + \alpha_D \log \left(\frac{T}{T_0} \right) \right] \end{cases} \quad (12)$$

Boundary conditions include an albedo boundary condition on $\partial\Omega$ in the axial direction and in the blanket salt and a homogeneous Neumann condition has been used for the concentration of the precursors

$$D \nabla \phi \cdot \mathbf{n} = -\gamma \phi, \quad \text{on } \partial\Omega. \quad (13)$$

The fluid flow is modelled as incompressible. Turbulence is treated using the realizable $k-\epsilon$ model with standard wall function treatment. The Boussinesq approximation has been used to consider buoyancy effects. Table 1 summarizes the quantities involved in Eqs. (11) to (13). Fig. 3 reports the numerical mesh used for the high-fidelity model: it counts 22671 hexahedral elements. The blank space represents the fuel blanket, not modelled here. The heat exchanger with the intermediate cycle is located right off the blanket. Dimensions are 2.05×2.26 m. The heat exchanger section is 0.2 m wide.

The transient considered in this work is a LOHS (Loss of Heat Sink) transient. Starting from nominal power 3 GW_{th} and critical conditions, the heat transfer coefficient between the reactor core and the heat sink is reduced by 25% with a constant mass flow rate. The high-fidelity transient simulation is carried out for 25 s with a time step of 0.001 s,

Table 1
MSFR equation nomenclature.

Symbol	Meaning
ϕ	Neutron flux.
v	Average neutron velocity.
β_t	Total delayed neutron fraction.
D	Diffusion coefficient.
Σ_a	Absorption macroscopic cross-section.
Σ_f	Fission macroscopic cross-section.
v	Average neutron produced per fission.
c_i	Delayed i th neutron precursor concentration.
λ_i	Decay constant of the i th precursor group.
γ	Albedo coefficient.
T	Salt temperature
ρ	Salt density
\mathbf{u}	Salt velocity.
ν_t	Fuel salt turbulent kinematic viscosity.
Sc_t	Turbulent Schmidt number.
d_j	i -th decay heat precursor concentration.
$\lambda_{h,j}$	Decay constant of the i -th decay heat precursor.
$\beta_{h,j}$	Fraction of the i -th decay heat precursor.
E_f	Average energy released per fission.
K_s	Laminar salt thermal conductivity.
K_t	Turbulent salt thermal conductivity.
C_p	Fuel specific heat.

and it required several computational hours on a 16-core machine of an HPC cluster.

4. Numerical results

This work focuses on three applications: the direct reconstruction of temperature through TR-GEIM, the indirect estimation of velocity, pressure, flux and precursors' concentration from temperature measurements and the prediction capabilities using different training sets of snapshots. Time $t(s)$ is considered the only parameter of the problem, belonging to the space $D = [0, 25]$; the snapshots are sampled every 0.025 s. The library of sensors for the temperature is represented by Gaussian linear functionals as in Introini et al. (2023b) and Riva et al. (2023a) with a given point-spread $s = 0.025$; three different sets of available positions will be tested: firstly, the TR-GEIM algorithm will be free to place the sensors everywhere in the domain (label: ID), then the selection will be restricted only on the whole boundary of the domain (label: BF), in the end, the left vertical boundary is excluded and only the remaining portions are considered (label: BP); the last two strategies mimic the real configuration in which in-core measurements are not allowed, therefore assessing the capabilities of inferring the quantities of interest using only out-core measurements. The techniques used in this work have been implemented in Python and are available in the `pyforce` package, built upon the FEniCSx library (Baratta et al., 2023; Scroggs et al., 2022a,b; Alnaes et al., 2014). Each application is characterized by the offline-online paradigm: to summarize, the former assesses the reducibility of the problem (this assessment is typically performed by looking at the eigenvalues of the POD per each field and making sure their decay is rather fast), generates the basis functions and selects the optimal positions of the sensors; the latter focuses on the estimation of the fields and the computation of errors (if the true field is known). In this work, errors are measured in the L^2 space as absolute and relative differences as follows

$$\begin{cases} E_M[u] = \|u(\mathbf{x}; \boldsymbol{\mu}) - \mathcal{P}_M[u](\mathbf{x}; \boldsymbol{\mu})\|_{L^2(\Omega)} = \|r(\mathbf{x}; \boldsymbol{\mu})\|_{L^2(\Omega)} \\ \varepsilon_M[u] = \frac{\|u(\mathbf{x}; \boldsymbol{\mu}) - \mathcal{P}_M[u](\mathbf{x}; \boldsymbol{\mu})\|_{L^2(\Omega)}}{\|u(\mathbf{x}; \boldsymbol{\mu})\|_{L^2(\Omega)}} = \frac{\|r(\mathbf{x}; \boldsymbol{\mu})\|_{L^2(\Omega)}}{\|u(\mathbf{x}; \boldsymbol{\mu})\|_{L^2(\Omega)}} \end{cases} \quad (14)$$

given $\mathcal{P}_M[u(\cdot; \boldsymbol{\mu})]$ a "projection" operator taking as an input the true field and returning the reconstruction using MOR with M basis functions. In the end, the measurements $\mathbf{y} \in \mathbb{R}^M$ are generated by the snapshots by computing the action of the functionals $\{v_m\}_{m=1}^M$ to the snapshots. The data of the online phase are polluted by random noise ϵ_m , modelled as a random variable distributed as a normal $\mathcal{N}(0, \sigma^2)$ in which $\sigma = 2.3$ K (Arora et al., 2021).

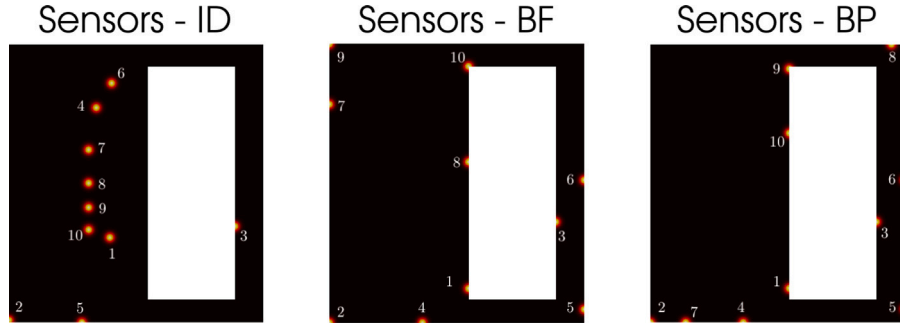


Fig. 4. GEIM sensors positioning for the free and boundary case.

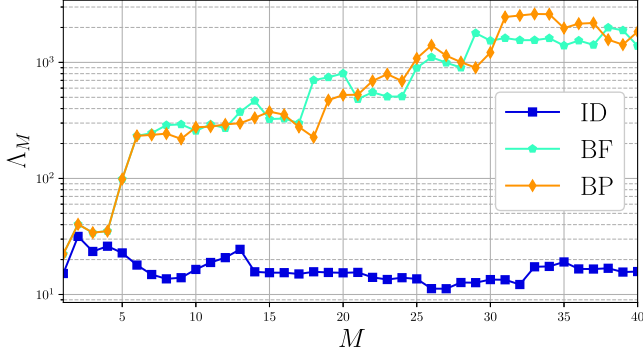


Fig. 5. Lebesgue Constant for each sensor configurations.

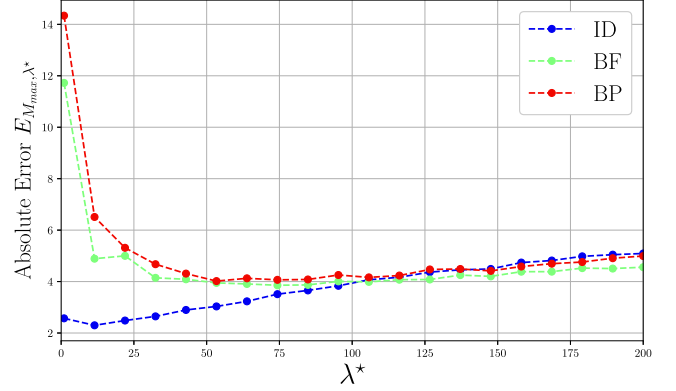


Fig. 6. Hyperparameter tuning for the different configurations through the absolute reconstruction error.

4.1. Direct reconstruction on temperature

The snapshots coming from the simulation are separated into a train and test set as follows

$$\begin{cases} \Xi_{\text{train}} = [0 : 0.025 : 20] \setminus [0.1 : 0.1 : 20] \\ \Xi_{\text{test}} = [0.1 : 0.1 : 20] \\ \Xi_{\text{predict}} = [20 : 0.025 : 25] \end{cases} \quad (15)$$

such that $\Xi_{\text{train}} \cap \Xi_{\text{test}} = \emptyset$.

Once the magic functions and magic sensors for the temperature, considering all three scenarios for sensors positions, have been generated, the temperature field can be reconstructed to compute the test errors at different values of the dimension of the reduced space. The locations of the sensors selected by the TR-GEIM in the three cases are shown in Fig. 4. As expected, without constraints TR-GEIM tends to select sensors where the temperature is the highest: constraining the problem to the core boundaries dramatically changes the sensors positioning, as the preferred optimal locations previously selected are no longer available. On the other hand, BF and BP do not show significant differences as seen in Fig. 4, the two sensors selected by the BF case on the inner boundary have rank 7 and 9 out of 10 sensors total, thus they have less importance than higher rank sensors whose optimal position is, instead, allowed in both configurations.

As mentioned in Section 2, the Lebesgue constant is an important quantity that can be used to assess *a priori* the reliability of the GEIM interpolation procedure. Fig. 5 shows how Λ_M grows as more sensors are added: the two constrained cases present a non-convergent behaviour compared to the unconstrained case, which means that the boundedness of the interpolation error is weaker the more constrained the problem is. As such, the regularization procedure must be calibrated for each case separately to ensure convergent behaviour, and thus boundedness of the interpolation error, for all cases under consideration.

As shown in Eq. (4), the linear system to be solved for the online phase of the TR-GEIM algorithm depends on the hyperparameter λ .

Table 2

Results of the hyperparameter tuning.

Configuration	ID	BF	BP
λ_{opt}^*	1	65.71	62.92

In Introini et al. (2023a), it is stated that the optimal value of this hyperparameter is proportional to the variance of the Gaussian random noise, i.e. $\lambda = \sigma^2$. However, the fact that the sensors positioning is constrained requires proper tuning of this parameter to achieve boundedness (as shown in Fig. 5). Therefore, the test set Ξ_{test} is further split into a validation set and a new test set (ratio 1:3) to find the optimal value of λ , expressed as the product of two terms λ^* and σ^2 : accordingly, the hyperparameter tuning problem reads

$$\lambda_{\text{opt}}^* = \underset{\lambda^* \in [1, 200]}{\operatorname{argmin}} E_{M_{\text{max}}, \lambda^*} = \underset{\lambda^* \in [1, 200]}{\operatorname{argmin}} \left\langle \left\| T(t) - I_{M_{\text{max}}}^{\lambda^*} [T](t) \right\|_{L^2(\Omega)} \right\rangle_{t \in \Xi_{\text{valid}}}, \quad (16)$$

given $\langle \cdot \rangle$ the sample mean operator and M_{max} the number of sensors used: the tuning process has been performed with $M_{\text{max}} = 25$. In practice, and the value of λ^* is chosen as the one that minimizes the absolute interpolation error considering the maximum number of sensors available.

The optimization problem has been solved in 2 steps: at first, the domain space of the hyperparameter has been discretized using 20 points in the interval $[1, 200]$ and for each one the average absolute error with M_{max} sensors is obtained and plotted in Fig. 6; then, the minimum of these 20 samples is computed and a second step of the optimization is performed starting from the guess value just obtained.

In the end, the results of the optimization are reported in Table 2. and they show that the constrained cases require a higher value, and thus a higher importance, of the regularization term, compared to

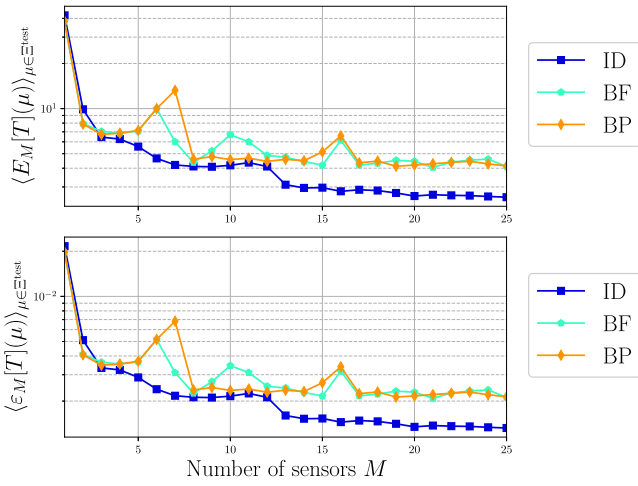


Fig. 7. Average test errors (absolute and relative) of the temperature field using TR-GEIM for the different sensor configurations.

the non-constrained case, confirming the behaviour of the Lebesgue constant.

Fig. 7 shows how the reconstruction of the temperature field behaves as the number of sensors increases using TR-GEIM. The optimal hyperparameter value found with the optimization procedure described above improves the reconstruction error: indeed, every configuration retrieves a good convergence trend, without instabilities due to noisy data (Argaud et al., 2017; Introini et al., 2023a). The positive effect of the tuning of λ^* heals the unboundedness of the Lebesgue constant (Fig. 5): in fact, the role of λ^* consists in weakening the interpolation condition of GEIM and, in this way, the Lebesgue constant becomes lower.

In Fig. 8, the contour plots of the reconstructed fields and the residual field, i.e. the absolute difference between the truth and the interpolant: not only the unconstrained ID case, but also the two constrained ones correctly reconstruct the temperature profile in the whole domain, making the direct reconstruction of fields in the whole domain of interest from out-core sparse noisy data feasible, since the difference between the three strategies for the sensor positioning is negligible.

4.2. Indirect reconstruction of unobservable fields

Considering the same train and test sets defined previously, the capability to indirectly reconstruct the unobservable fields characterizing the system will be tested. During the offline phase, POD modes for the required fields (velocity, neutron flux and precursor group 6 – as an example –) are generated, and the mapping functions $F_n(t) = \alpha_n(t)$ are created. For this case, cubic piecewise interpolation is adopted both for the GEIM and the POD coefficients for the other fields.

The estimation phase to find the parameter from the measurements of the temperature has been performed for all the strategies of sensors' selection (i.e., ID, BF and BP), giving comparable results, as seen in see Fig. 9. This phase is the most challenging one, because of the global optimization problem that must be solved during the online step: this may be an issue for highly non-convex cost functions since the time required to reach the minimum can be high (whereas the online phase should be kept as computationally cheap as possible). In this case, in which there is only one parameter, the solution is almost instantaneous.

Furthermore, a fundamental underlying hypothesis of this version of the IR algorithm is the bijectivity of the map $t \rightarrow T(t)$ (Introini et al., 2023b; Riva et al., 2023a) and in this case this hypothesis is not always valid: this implies that for different values of t there are states $T(t)$ very similar to one another. This could result in a failure for

the PE phase, as shown in Fig. 10 which reports the Q-Q plot of the parameter estimation for the different sensor configurations: after 10 s, the optimization fails to converge, and the parameter is not accurately estimated. This is because a steady state is reached, and all the states become very similar to one another.

Nevertheless, it is important to analyse the reconstruction error of unobservable fields with all the strategies for sensor placement. Fig. 11 reports the relative reconstruction error of the POD-I using the parameter estimated with 20 measurements with varying number of POD modes: for every field the reconstruction is quite accurate reaching average relative errors lower than 0.5%. Moreover, in any case a saturation point is reached making the addition of more modes useless, as in Introini et al. (2023b), this is due to the fact that the parameter (i.e., the time) can be estimated up to certain accuracy and hence the modal coefficients of the POD cannot be perfectly estimated resulting in a saturation effect, so that adding more basis does not have beneficial effects on the reconstruction.

Figs. 12, 13 and 14 show the glyph and contour plots of the velocity, the neutron flux and the sixth precursors group at $t = 19.90$ s: every case the reconstructed field is very close to the true one and the difference is negligible. Therefore, the POD-I can accurately reconstruct all the fields, given a good estimation of the parameter. The most important limitations of this method lay in the parameter estimation phase and the correct choice of the reduced space for the indirectly reconstructed fields: at the moment, the optimization and the linear subspace spanned by POD modes are sufficient to accurately predict the behaviour of the quantities of interest, however, it will be a matter of future studies to use advanced MOR methods with non-linear reduction (Phillips et al., 2021) and to integrate machine learning algorithms for the parameter estimation (Phillips et al., 2021; Rasmussen and Williams, 2006).

4.3. Prediction capabilities of (TR-)GEIM

In this Section, the extrapolation/prediction capabilities of TR-GEIM are tested, by using snapshots with correspondent parameter outside the training domain. In particular, the parameters are sampled in Ξ_{predict} . The magic functions and magic sensors obtained from the previous sections are used, along with the optimal values of the hyperparameter λ^* . The prediction capabilities of TR-GEIM will be assessed using noisy data, polluted with the same variance of the previous sections.

Fig. 15 shows the average absolute and relative reconstruction error of the algorithm for the different sensor configurations: each case returns a good estimation of the state of the system showing that the TR-GEIM algorithm is suited for prediction. Nevertheless, the errors reach a saturation point, thus as more sensors are added the prediction does not improve globally, because of the inevitable numerical errors of the MOR methods, moreover the data are taken outside the training space and hence they represent unseen behaviour showing that the training information can reach up to a certain level of accuracy. To better correct the mathematical model, especially near the sensors, other MOR techniques may be used such as the Parametrized-Background Data-Weak (PBDW) formulation (Maday et al., 2014), generalizing the theory underlying Hybrid Data Assimilation methods. This work adopts a fully-coupled MP model as the *ground-truth* describing the physics of the reactor, and it is used both for training and testing of the data-driven MOR techniques. However, this approach strongly depends by the underlying hypothesis and approximation made in the development of the coupling between neutronics and thermal-hydraulics; conversely, the PBDW formulation describes a general framework for data-driven MOR techniques able to update/correct the information of the mathematical modelling, as done by the authors in Riva et al. (2024).

In the end, the temperature is plotted at the end of the transient, $t = 25$ s, in Fig. 16: the prediction is quite accurate, highlighting again the reliability of TR-GEIM even in the extrapolation context.

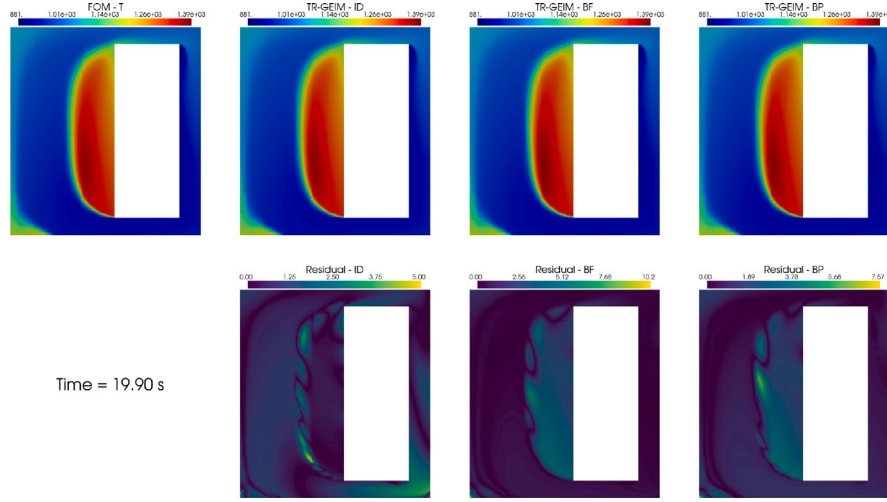


Fig. 8. Comparison of the reconstructed and residual fields, using contour plots, for the different sensor configurations at time $t = 19.90$ s with measurements polluted by random noise with $\sigma = 2.3$ K.

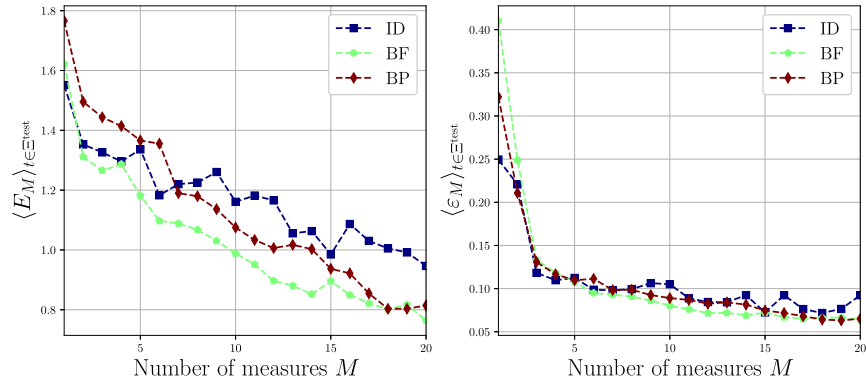


Fig. 9. Absolute and Relative error on the parameter obtained by the Parameter Estimation phase using varying number of measurements.

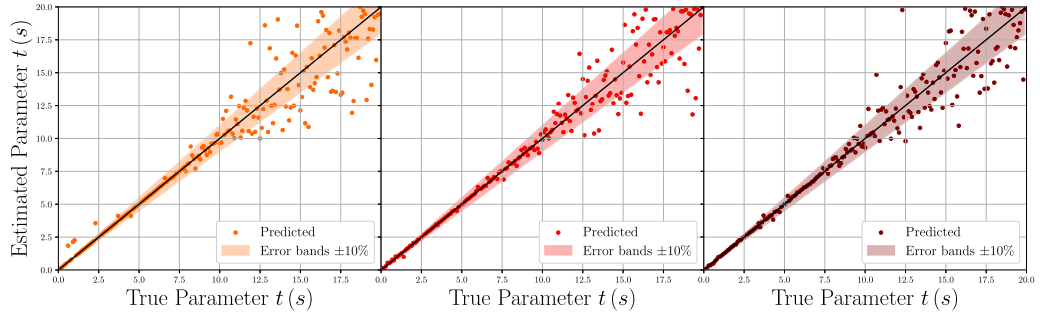


Fig. 10. Q-Q plot of the PE phase for the different sensor configurations (from left to right ID, BF, BP) using 20 measurements.

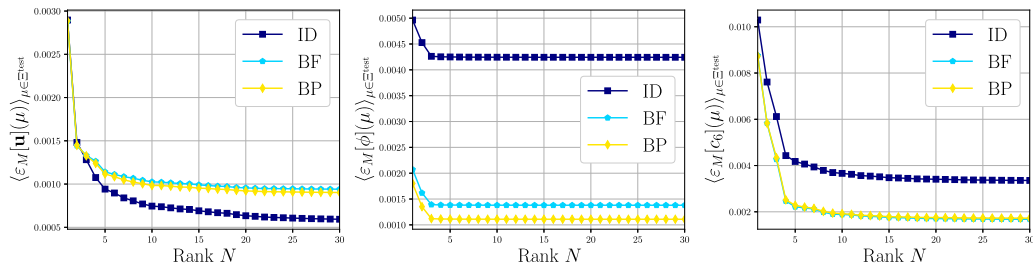


Fig. 11. Average test errors (relative) of the other fields field using POD-I, with the parameter estimated using all the sensor placement strategies.

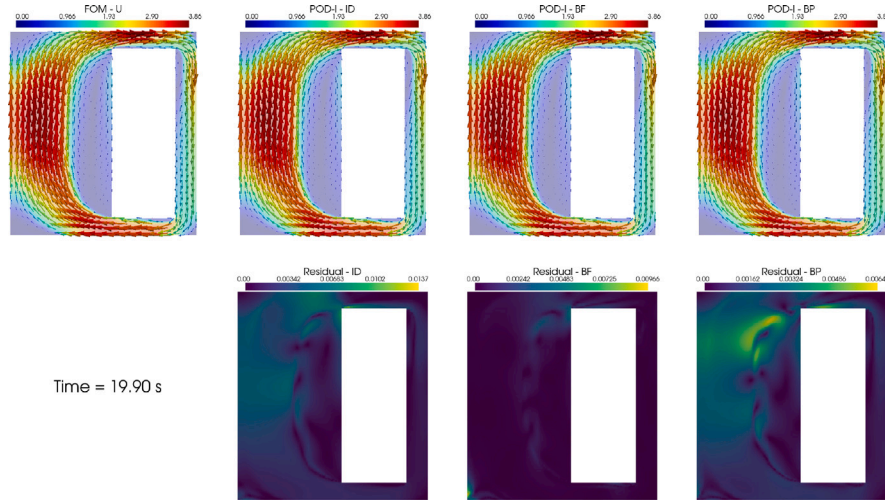


Fig. 12. Glyph plots of the velocity field and the POD-I reconstructions and the residual field for each different configuration at $t = 19.90$ s.

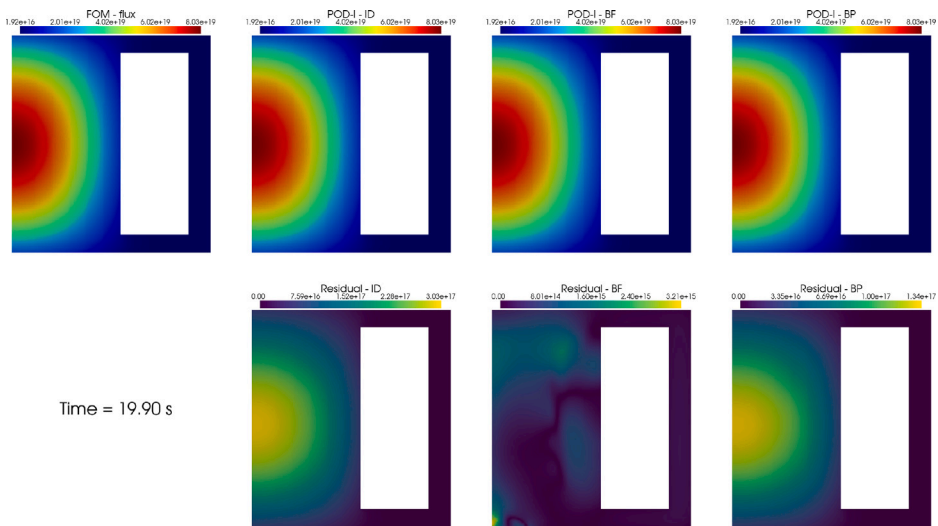


Fig. 13. Contour plots of the neutron flux field and the POD-I reconstructions and the residual field for each different configuration at $t = 19.90$ s.

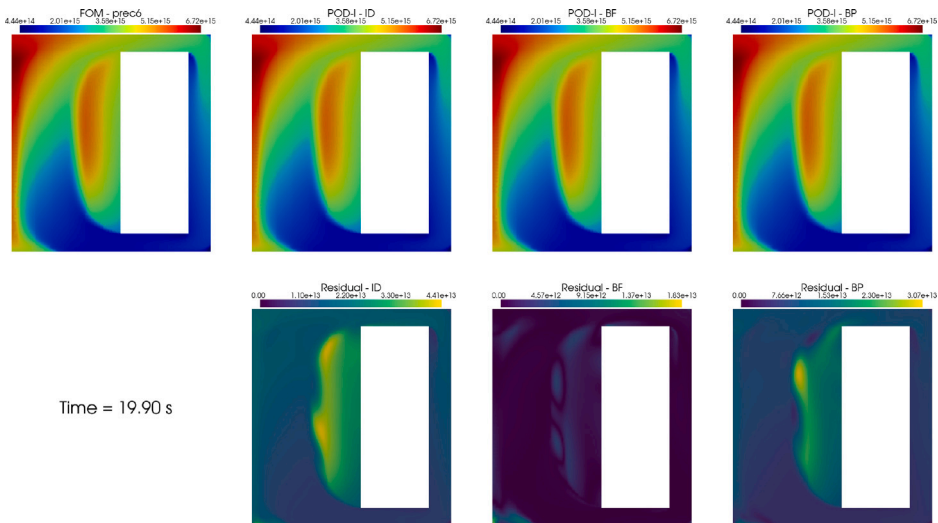


Fig. 14. Contour plots of the 6-th group of precursors field and the POD-I reconstructions and the residual field for each different configuration at $t = 19.90$ s.

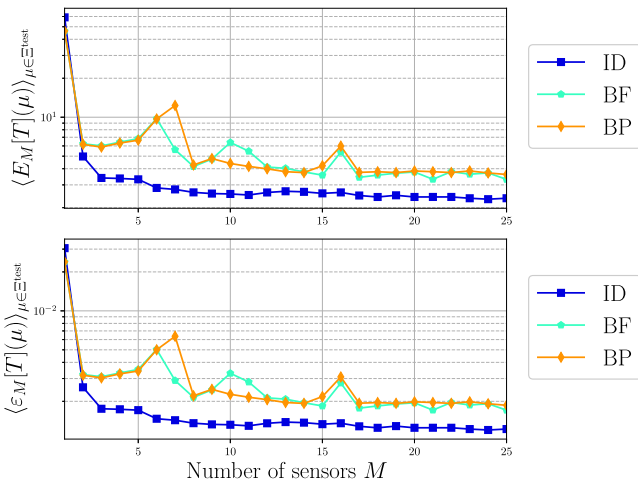


Fig. 15. Average test error for the prediction phase using different sensor configurations.

4.4. Computational cost of the online phase

The MOR techniques have been conceived to reduce the computational burden of the numerical solution of mathematical models described by PDEs. More in general, for what concerns the computational resources necessary to implement data-driven MOR techniques: the generation of the snapshots is typically done on High-Performance Computing (HPC) clusters since the FOM must be solved several times to get the training dataset; the offline phase, albeit still computationally demanding, is performed only once; conversely, the online step can be very cheap, as solving the low-dimensional TR-GEIM linear system is straightforward, and can be easily performed on personal computers. The bottleneck for the online phase is represented by the solution of the optimization problem (8), especially in the presence of several local minima. For this reason, the authors are currently investigating other methods to overcome this limitation.

The results of the online phase have been obtained using a quad-core Intel Core i5 processor (1.4 GHz) installed on a personal computer: collecting the data (computing the action of the magic sensors), solving for the TR-GEIM matrix, whose size is at most 25×25 , and reconstructed the whole field, including error computations, have taken around 0.03 s, almost constant with respect to the number of measurements; whereas the parameter estimation phase is linearly dependent on the number of measurements, ranging from 0.1 to 0.3 s for each solve.

5. Conclusions

In this work, the problem of reconstructing the entire state of a system starting from partial measurements has been studied using the Generalised Empirical Interpolation Method, regularized with Tikhonov in presence of noisy data, and an Indirect Reconstruction algorithm: firstly, the measurements are used to estimate the measured field and the characterizing parameters of the system, then the Proper Orthogonal Decomposition with Interpolation is used to reconstruct all the unobservable fields. Moreover, the prediction capabilities of the TR-GEIM algorithm have been discussed. This work has focused on the possibility of inferring and predicting the full state of a CFR (2D EVOL geometry), either considering available sensors inside the domain or placed only on the whole or partial boundary.

The TR-GEIM algorithm is a robust technique to reconstruct the temperature field, in fact, both in-core and out-core measurements of it enable a correct estimation of the overall field, reaching relative errors measured in L^2 around 0.5% with 25 measures. Good results are also

obtained for the Indirect Reconstruction of non-observable fields, such as velocity, pressure, neutron flux and precursors concentration, even though the parameter cannot be accurately estimate when the bijectivity of the map $\mu \rightarrow T(\mu)$ is lost. In the end, the prediction capabilities of TR-GEIM have been tested: each strategy produces comparable and accurate results, making TR-GEIM suited for extrapolation problems.

It will be a matter of future studies to use other sensor positioning algorithms, such as the SGREEDY algorithm (Maday et al., 2014). Furthermore, it is foreseen to adopt Machine Learning techniques to improve and speed up the parameter estimation phase for the Indirect Reconstruction algorithm. These techniques will also be tested against different reactors configurations and types, and thanks to their generality can be extended for the analysis of non-nuclear systems as well (for example, to study hybrid energy systems or in the design of experimental facilities as they offer an a-priori estimation of the optimal positions for the instrumentation).

List of symbols

Acronyms

BF	Full Boundary case for sensor placement
BP	Partial Boundary case for sensor placement
CFR	Circulating Fuel Reactor
DA	Data Assimilation
FOM	Full Order Model
GEIM	Generalized Empirical Interpolation Method
ID	Ideal case for sensor placement
HPC	High Performance Computing
IR	Indirect Reconstruction
MOR	Model Order Reduction
MSFR	Molten Salf Fast Reactor
ODE	Ordinary Differential Equation
PBDW	Parameterized-Background Data-Weak
PE	Parameter Estimation phase
POD	Proper Orthogonal Decomposition
POD-I	Proper Orthogonal Decomposition with Interpolation
PDE	Partial Differential Equation
TR-GEIM	Tikhonov Regularization Generalized Empirical Interpolation Method

Greek Symbols

α_m	m -th POD reduced coefficient
β_m	m -th GEIM reduced coefficient
ϵ	Random Noise
λ	Regularization parameter
ψ_m	m -th POD mode
σ	Standard deviation of noise and of the training GEIM coefficient
ϵ_M	Relative Error using M measurement/basis functions
Λ	Lebesgue constant
Ω	Spatial domain
Υ	Library of sensors
Ξ	Subset of the parameter space
μ	Vector of parameters

Latin symbols

q_m	m -th magic function
r	Residual field, i.e. absolute difference between true field and reconstruction
s	Point spread of the sensor
u	Generic Function
v	Functional representing the sensor
y_m	m -th measurement
E_M	Absolute Error using M measurement/basis functions
J	Loss function of the parameter estimation phase
M	Number of sensors/measurements/basis functions
T	Temperature

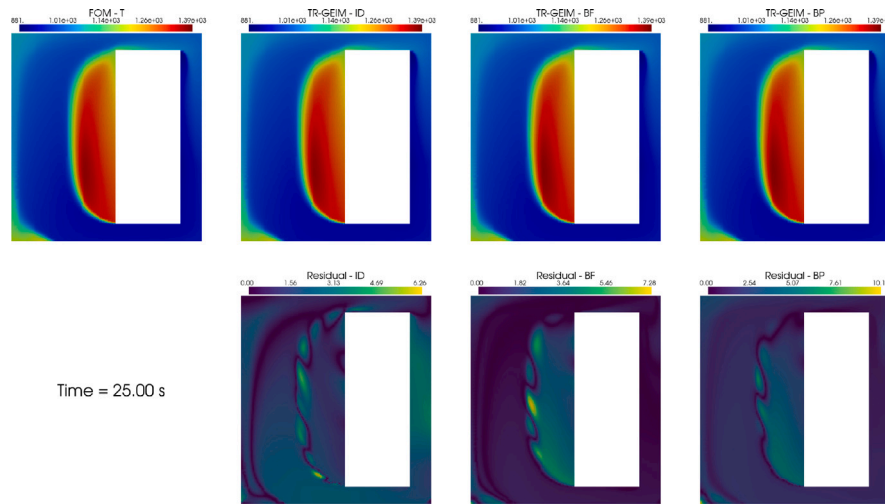


Fig. 16. Contour plots of the temperature field and the POD-I reconstructions and the residual field for each different configurations at $t = 25$ s.

\mathbb{B}	GEIM matrix
\mathbb{R}	Set of real numbers
\mathbf{T}	Tikhonov regularization matrix
\mathcal{D}	Parameter space
\mathcal{F}	Map to the GEIM coefficient
\mathcal{I}	GEIM interpolant
\mathcal{N}	Gaussian normal distribution
\mathcal{P}_M	Projection operator using M measurements/basis functions
\mathcal{U}	Hilbert space of functions
\mathbf{x}	Spatial vector

CRedit authorship contribution statement

Antonio Cammi: Writing – review & editing, Supervision, Formal analysis. **Stefano Riva:** Writing – original draft, Software, Investigation, Data curation. **Carolina Introini:** Writing – review & editing, Methodology, Conceptualization. **Lorenzo Loi:** Writing – original draft, Visualization. **Enrico Padovani:** Writing – review & editing, Supervision.

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.

Data availability

The **pyforce** package used in this work is part of the ROSE framework (Reduced Order multi-phys data-drivEn) developed by the authors at the **ERMETE Lab**. The package can be installed using **pip** and will be soon available on Github at <https://github.com/ERMETE-Lab/ROSE-pyforce> under the MIT license.

References

- Alnaes, M.S., Logg, A., Ølgaard, K.B., Rognes, M.E., Wells, G.N., 2014. Unified form language: A domain-specific language for weak formulations of partial differential equations. *ACM Trans. Math. Softw.* 40, <http://dx.doi.org/10.1145/2566630>.
- Argaud, J.-P., Bouriquet, B., de Caso, F., Gong, H., Maday, Y., Mula, O., 2018. Sensor placement in nuclear reactors based on the generalized empirical interpolation method. *J. Comput. Phys.* 363, 354–370. <http://dx.doi.org/10.1016/j.jcp.2018.02.050>, URL <https://www.sciencedirect.com/science/article/pii/S0021999118301414>.

- Argaud, J.-P., Bouriquet, B., Gong, H., Maday, Y., Mula, O., 2017. Stabilization of (G)EIM in presence of measurement noise: Application to nuclear reactor physics. In: Bittencourt, M.L., Dumont, N.A., Hesthaven, J.S. (Eds.), *Spectral and High Order Methods for Partial Differential Equations ICOSAHOM 2016*. Springer International Publishing, Cham, pp. 133–145.
- Arora, O., Lancaster, B., Yang, S.R., Vaghetto, R., Hassan, Y., 2021. Advanced flow and temperature measurements in a forced convection molten salt test loop. *Ann. Nucl. Energy* 159, 108269. <http://dx.doi.org/10.1016/j.anucene.2021.108269>, URL <https://www.sciencedirect.com/science/article/pii/S0306454921001456>.
- Aufiero, M., Cammi, A., Geoffroy, O., Losa, M., Luzzi, L., Ricotti, M.E., Rouch, H., 2014. Development of an OpenFOAM model for the Molten Salt Fast Reactor transient analysis. *Chem. Eng. Sci.* 111, 390–401. <http://dx.doi.org/10.1016/j.ces.2014.03.003>, URL <https://www.sciencedirect.com/science/article/pii/S0009250914001146>.
- Baratta, I.A., Dean, J.P., Dokken, J.S., Habera, M., Hale, J.S., Richardson, C.N., Rognes, M.E., Scroggs, M.W., Sime, N., Wells, G.N., 2023. DOLFINx: the next generation FEniCS problem solving environment. <http://dx.doi.org/10.5281/zenodo.10447666>, preprint.
- Brunton, S.L., Kutz, J.N., 2022. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, second ed. Cambridge University Press, USA.
- Cao, P., Cao, C., Gan, Q., 2021. A 3-D neutron distribution reconstruction method based on the off-situ measurement for reactor. *IEEE Trans. Nucl. Sci.* 68 (12), 2694–2701. <http://dx.doi.org/10.1109/TNS.2021.3123381>.
- Carrasi, A., Bocquet, M., Bertino, L., Evensen, G., 2018. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *WIREs Clim. Chang.* 9 (5), e535. <http://dx.doi.org/10.1002/wcc.535>.
- Daescu, D.N., Navon, I.M., 2007. Efficiency of a POD-based reduced second-order adjoint model in 4D-var data assimilation. *Internat. J. Numer. Methods Fluids* 53 (6), 985–1004. <http://dx.doi.org/10.1002/fld.1316>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/fld.1316>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/fld.1316>.
- Demo, N., Tezzele, M., Mola, A., Rozza, G., 2019. A complete data-driven framework for the efficient solution of parametric shape design and optimisation in naval engineering problems. In: *MARINE 2019: VIII International Conference on Computational Methods in Marine Engineering*. pp. 1–12.
- Gong, H., Argaud, J.-P., Bouriquet, B., Maday, Y., 2016. The empirical interpolation method applied to the neutron diffusion equations with parameter dependence. In: *Physics of Reactors 2016, PHYSOR 2016: Unifying Theory and Experiments in the 21st Century*, vol. 1, (no. May), pp. 54–63.
- Gong, H., Chen, Z., Maday, Y., Li, Q., 2021. Optimal and fast field reconstruction with reduced basis and limited observations: Application to reactor core online monitoring. *Nucl. Eng. Des.* 377, 111113. <http://dx.doi.org/10.1016/j.nucengdes.2021.111113>, URL <https://www.sciencedirect.com/science/article/pii/S0029549321000650>.
- Gong, H., Cheng, S., Chen, Z., Li, Q., 2022. Data-enabled physics-informed machine learning for reduced-order modeling digital twin: Application to nuclear reactor physics. *Nucl. Sci. Eng.* 196 (6), 668–693. <http://dx.doi.org/10.1080/00295639.2021.2014752>, Publisher: Taylor & Francis.
- Introini, C., 2021. *Advanced Modelling and Stability Analysis for Nuclear Reactors* (Ph.D. thesis). Politecnico di Milano.
- Introini, C., Cavalleri, S., Lorenzi, S., Riva, S., Cammi, A., 2023a. Stabilization of Generalized Empirical Interpolation Method (GEIM) in presence of noise: A novel approach based on Tikhonov regularization. *Comput. Methods Appl. Mech. Engrg.* 404, 115773. <http://dx.doi.org/10.1016/j.cma.2022.115773>, URL <https://www.sciencedirect.com/science/article/pii/S0045782522007290>.

- Introini, C., Riva, S., Lorenzi, S., Cavalleri, S., Cammi, A., 2023b. Non-intrusive system state reconstruction from indirect measurements: A novel approach based on Hybrid Data Assimilation methods. *Ann. Nucl. Energy* 182, 109538. <http://dx.doi.org/10.1016/j.anucene.2022.109538>, URL <https://www.sciencedirect.com/science/article/pii/S0306454922005680>.
- Lassila, T., Manzoni, A., Quarteroni, A., Rozza, G., 2014. Model order reduction in fluid dynamics: Challenges and perspectives. In: *Reduced Order Methods for Modeling and Computational Reduction*. Springer International Publishing, Cham, pp. 235–273. http://dx.doi.org/10.1007/978-3-319-02090-7_9.
- Maday, Y., 2006. Reduced basis method for the rapid and reliable solution of partial differential equations. In: *International Congress of Mathematicians*, Vol. 3. ICM 2006, pp. 1255–1270. <http://dx.doi.org/10.4171/022-3/60>.
- Maday, Y., Mula, O., 2013. A generalized empirical interpolation method: application of reduced basis techniques to data assimilation. In: *Springer INdAM Series*, Springer Milan, ISBN: 9788847025929, pp. 221–235. http://dx.doi.org/10.1007/978-88-470-2592-9_13.
- Maday, Y., Mula, O., Patera, A.T., Yano, M., 2015. The generalized empirical interpolation method: Stability theory on Hilbert spaces with an application to the Stokes equation. *Comput. Methods Appl. Mech. Engrg.* 287, 310–334. <http://dx.doi.org/10.1016/j.cma.2015.01.018>, Publisher: Elsevier B.V.
- Maday, Y., Mula, O., Turinici, G., 2016. Convergence analysis of the generalized empirical interpolation method. *SIAM J. Numer. Anal.* 54 (3), 1713–1731. <http://dx.doi.org/10.1137/140978843>, Publisher: Society for Industrial & Applied Mathematics (SIAM).
- Maday, Y., Patera, A., Penn, J., Yano, M., 2014. A parameterized-background data-weak approach to variational data assimilation: formulation, analysis, and application to acoustics. *Internat. J. Numer. Methods Engrg.* 102, <http://dx.doi.org/10.1002/nme.4747>.
- Pei, C., Quan, G., Jing, S., Pengcheng, L., Fang, W., Liqin, H., Yican, W., 2020. An artificial neural network based neutron field reconstruction method for reactor. *Ann. Nucl. Energy* 138, 107195. <http://dx.doi.org/10.1016/j.anucene.2019.107195>, URL <https://www.sciencedirect.com/science/article/pii/S0306454919307054>.
- Phillips, T.R.F., Heaney, C.E., Smith, P.N., Pain, C.C., 2021. An autoencoder-based reduced-order model for eigenvalue problems with application to neutron diffusion. *Internat. J. Numer. Methods Engrg.* 122 (15), 3780–3811. <http://dx.doi.org/10.1002/nme.6681>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.6681>, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.6681>.
- Quarteroni, A., Manzoni, A., Negri, F., 2015. *Reduced basis methods for partial differential equations: An introduction*, first ed. UNITEXT, Springer Cham, <http://dx.doi.org/10.1007/978-3-319-15431-2>.
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian processes for machine learning*. In: *Adaptive Computation and Machine Learning*, MIT Press.
- Riva, S., Introini, C., Cammi, A., 2024. Multi-physics model bias correction with data-driven reduced order modelling techniques: Application to nuclear case studies. [arXiv:2401.07300](https://arxiv.org/abs/2401.07300).
- Riva, S., Introini, C., Lorenzi, S., Cammi, A., 2023a. Hybrid data assimilation methods, part I: Numerical comparison between GEIM and PBDW. *Ann. Nucl. Energy* 190, 109864. <http://dx.doi.org/10.1016/j.anucene.2023.109864>, URL <https://www.sciencedirect.com/science/article/pii/S0306454923001834>.
- Riva, S., Introini, C., Lorenzi, S., Cammi, A., 2023b. Hybrid Data Assimilation methods, Part II: Application to the DYNASTY experimental facility. *Ann. Nucl. Energy* 190, 109863. <http://dx.doi.org/10.1016/j.anucene.2023.109863>, URL <https://www.sciencedirect.com/science/article/pii/S0306454923001822>.
- Rozza, G., Hess, M., Stabile, G., Tezzele, M., Ballarin, F., Gräßle, C., Hinze, M., Volkwein, S., Chinesta, F., Ladeveze, P., Maday, Y., Patera, A., Farhat Char, J., 2020. *Model Order Reduction: Volume 2: Snapshot-Based Methods and Algorithms*. De Gruyter, <http://dx.doi.org/10.1515/9783110671490>.
- Scroggs, M.W., Baratta, I.A., Richardson, C.N., Wells, G.N., 2022a. Basix: a runtime finite element basis evaluation library. *J. Open Source Softw.* 7 (73), 3982. <http://dx.doi.org/10.21105/joss.03982>.
- Scroggs, M.W., Dokken, J.S., Richardson, C.N., Wells, G.N., 2022b. Construction of arbitrary order finite element degree-of-freedom maps on polygonal and polyhedral cell meshes. *ACM Trans. Math. Software* 48 (2), 18:1–18:23. <http://dx.doi.org/10.1145/3524456>.
- Serp, J., Allibert, M., Beneš, O., Delpech, S., Feynberg, O., Ghetta, V., Heuer, D., Holcomb, D., Ignatiev, V., Kloosterman, J.L., Luzzi, L., Merle-Lucotte, E., Uhlíř, J., Yoshioka, R., Zhimin, D., 2014. The molten salt reactor (MSR) in generation IV: Overview and perspectives. *Prog. Nucl. Energy* 77, 308–319. <http://dx.doi.org/10.1016/j.pnucene.2014.02.014>, URL <https://www.sciencedirect.com/science/article/pii/S0149197014000456>.