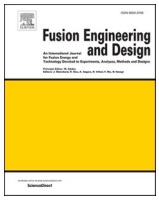
Fusion Engineering and Design 204 (2024) 114491

Contents lists available at ScienceDirect

Fusion Engineering and Design

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



A reduced-order model to estimate first wall particle and heat fluxes for systems codes

Tiago Pomella Lobo *, Sergey Pestchanyi, Ivan Alessio Maione

Karlsruhe Institute of Technology, Karlsruhe, Germany

ARTICLE INFO

Keywords:
Plasma filament
Far scrape-off layer (far-SOL)
Reduced order model (ROM)
TOKES
EU-DEMO
Systems code

ABSTRACT

Current systems codes (SCs) do not assess the impact of plasma filament parameters and dynamics in the SOL when performing systems-level analyses. Simulation tools based on simplified turbulent transport models could address such need, such as the TOKES code, but these tend to run in timescales that are prohibitive for incorporation in SCs. This work presents a technique to build surrogate models for SCs, obtained with the modes of Reduced-Order Models (ROMs), and is exemplified with results from TOKES. These surrogates were developed with Principal Components Analysis (PCA), multivariate regression against TOKES inputs (filament temperatures, densities, and ejection speeds) and k-fold cross-validation, applied to the results after transformations based on rational powers of the inputs to optimize the prediction capabilities of the surrogates. Selection was performed with a modified Kullback-Leibler Divergence (KLD), and validation, with withheld cases. Main results include validation of the methodology due to the relatively low number of modes needed to represent more than 90% of the data variance, and the data transformation exponents that optimize the regression.

1. Introduction

While most particles ejected from confined tokamak plasmas are directed towards the Divertor, some are distributed along the First Wall (FW) of Breeding Blankets (BBs) by following magnetic field lines in the so-called "far" Scrape-Off Layer (far-SOL, as opposed to the first, the near-SOL) [1]. Even though these fluxes are not the only phenomenon adding to the heat deposited on the FW (*e.g.* photons, charge-exchange), they can contribute to 20%−30% [2,3] of the challenging design issue that is the total heat expected on the FW armor (at least up to ~1 MW/m^2 in the EU-DEMO [4], not considering off-normal events). At the same time, these fluxes also contribute to particle implantation in the armor (ion bombardment), which can impact systems-level analyses such as tritium permeation into the FW coolant, tritium retention in BB steel, and effusion fluxes that lengthen the pump-down [5].

Thus, Fusion power plant systems must be designed with these heat and particle fluxes in mind, ideally in the form of *poloidal flux profiles* that depict pertinent information, *e.g.* "hot spots" on the FW. Empirical laws similar to the Eich scaling [6,7] are often employed for quickly estimating these profiles, which distribute the heat to be delivered to the FW over the flux lines in the SOL with an exponential expression characterized by a decay length, particular to each machine [8]. However, scaling laws are generally difficult to be modelled and validated for the far-SOL [9]. And, even when available, they do not compute consistent particle and heat fluxes, arguably of greater import in systemic analyses, since they do not model the processes with which these fluxes reach the walls.

Current tokamak plasma physics prescribes that the continuous production and expulsion of plasma filaments from the confined region is a main phenomenon contributing to these far-SOL fluxes. Filaments are plasma objects spanned along the magnetic field lines with cross-sections in the order of centimeters, that arise from small plasma fluctuations near the separatrix due to turbulence in that region. These plasma tubes propagate from the confined region into and across the SOL due to $\vec{E} \times \vec{B}$ drift, deforming along the way. However, they exhibit properties closer to the confined plasma from where they originate, instead of the SOL, which affects their dynamics and impact to the walls. Yet, a complete model of the process by which they are created and ejected remains a challenge for plasma physicists [10].

Nonetheless, the dynamics of filaments in the SOL can be approximated with simplified fluid-dynamics models as a palliative measure. The TOKES code [11] has been augmented to compute the evolution of plasma parameters inside filaments with prescribed initial cross-section diameters and ejection speeds, for fixed maps of the 2D poloidal magnetic flux and chamber wall that characterize a reactor design. The plasma in a filament starts with uniform temperature and density distributions along the magnetic field line found at the filament formation

* Corresponding author.
E-mail address: t.pomella-lobo@outlook.com (T. Pomella Lobo).

https://doi.org/10.1016/j.fusengdes.2024.114491
Received 13 October 2023; Received in revised form 3 May 2024; Accepted 4 May 2024
Available online 19 May 2024

position, between the pedestal and the separatrix. Its evolution is computed by coupling the plasma dynamics in the directions parallel and perpendicular to the field lines, which results in variations of plasma parameters along the magnetic field, and in flow of that plasma towards the wall. Renormalization of the deposited heat flux distribution on the wall by a prescribed far-SOL fraction of the transport power across the separatrix allows TOKES to compute the highly peaked total heat load profiles on the FW, alongside the associated (*i.e.* consistent) particle loads. Some peaks that can be up to 3 times higher than those underestimated by usual scaling laws, *e.g.* 60 kWm^{-2} instead of 20 near the upper X-point for a prescribed a far-SOL fraction of 69MW [12].

This type of result is a desirable addition to the MIRA systems code (SC) [13], which currently does not boast of any far-SOL transport modeling [14] and is under development to enable studies of the impact of a fixed reactor design on other power plant systems, such as in the Balance-of-Plant and the Fuel Cycle. It would provide high-fidelity load-consistent heat and particle flux profiles that can be used in the aforementioned systems-level analyses, while also being model-consistent with plasma transport/magnetic models, since MIRA already produces magnetic flux and wall maps, as well as 1D profiles for plasma parameters along the normalized flux coordinate. Subsequently, further model-consistency can be achieved between heat/particle loads from all SOL models with energy and mass balances and for a choice of partitioning between far-SOL, near-SOL and radiation contributions based on literature [2,3].

At the same time, direct coupling of TOKES to a SC is challenging; the timescale required for its calculations is not compatible with the framework of SCs, and its spatial meshing procedure is prone to require manual adjustment. That is, studies on the impact of these types of results in MIRA require employing an alternate approach. A potential strategy is the production of surrogate models with data-driven techniques and, of those, the Principal Components Analysis (PCA) [15] is a promising candidate because it is less training-data greedy than options such as Artificial Neural Networks (ANNs) [16–18], given the computational cost of TOKES runs. The application of such technique requires the construction of a database of results for a fixed reactor configuration, *i.e.* a specific plasma equilibrium and wall geometry, which antecedes the production of surrogates. This methodology may limit the flexibility of this kind of model in the framework of systems codes, but the approach can be considered sufficient for the desired systemic analyses described; at least while direct coupling with TOKES is not enabled by code refactoring and optimization.

This work presents a novel technique to build optimized surrogate models using the PCA, and the potential parametric dependencies it may reveal when applied to results of plasma simulations. It is exemplified with a collection of cases computed by the TOKES code for flux and wall maps representative of the European Demonstration Powerplant (EU-DEMO) and combinations of initial filament parameters (plasma temperature and density, and ejection speed). Section 2 describes the TOKES simulation results and the proposed methodology. Section 3 describes the PCA solution and subsequent surrogate models built by regression against filament parameters. Section 4 describes the optimization process, including a novel expression derived for model selection. Section 5 discusses the results, including the new parametric dependency identified and proposes follow-up studies.

2. Overview of the TOKES library and of the proposed methodology

A collection of cases was built by running the TOKES code with the poloidal flux and wall maps shown in Fig. 1, produced by MIRA, and for the 32 combinations of the following initial filament parameters,[1]

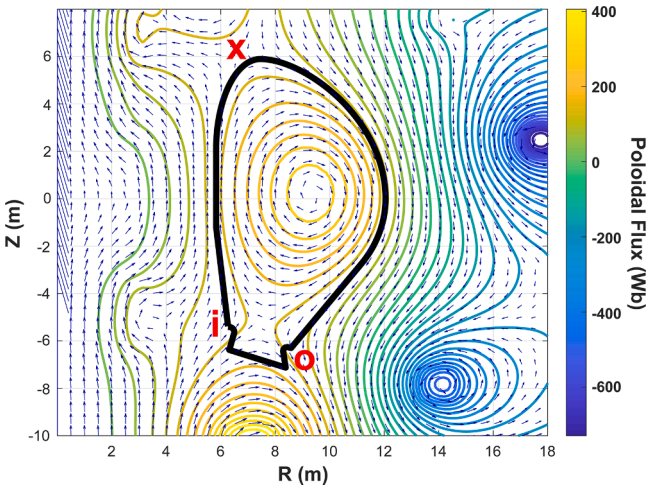


Fig. 1. Poloidal flux and reactor chamber wall maps computed by MIRA for the EU-DEMO and passed to TOKES as reactor configuration. Letter labels (red) indicate points of interest on the surface for which particle and heat flux profiles due to filaments are computed: lowest inboard (*i*) and outboard (*o*) points; the upper (ex-vessel) X-point is also indicated in the magnetic saddle point (*x*).

representatives for the reactor of interest:

- temperature $T(eV) \in [100,\ 300,\ 1000,\ 3000]$,
- density $N(m^{-3}) \in [10^{19},\ 10^{20}]$,
- ejection speed $V\left(\frac{m}{s}\right) \in [100,\ 200,\ 500,\ 1000]$.

Each case produced by a TOKES simulation, lasting 10–30 min of computational time, provided poloidal profiles for both the heat flux $\vec{Q}$ (MWm^{-2}) and the particle flux $\vec{G}$ $(m^{-2}s^{-1})$ reaching the FW, as 300-point vectors of values along the coordinate following the plasma chamber wall in counterclockwise direction, starting at the lowest inboard (*i-label*) point and through the lowest outboard (*o-label*) point.

Fig. 2 shows these profiles for a subset of the cases, normalized to 1 MW of prescribed far-SOL power. Miniaturized versions of the full profiles show the difference in scale between local maxima close to the Divertor (lowest inboard/outboard points) and the most significant local maxima seen near the upper X-point (*x label*). Flux values between the lowest inboard and outboard points are null for all cases because filaments in TOKES are restricted to the far-SOL and do not hit the Divertor. The red boxes in the miniaturized profiles depict the partial ranges in the profile plots that are magnified and displayed in the center of the figure; the different scales of these ranges, in comparison to the local maxima close to the Divertor, emphasizes the need of a model that concomitantly computes both heat and particle profiles, since one cannot be trivially derived from the other.

A confront between the magnified cases, and between magnified and miniaturized profiles of each case, also demonstrates the non-linearity of the data, which implies that estimating these profiles cannot be achieved with simple multiplication factors applied to a fixed "template" profile. That is, the impact that these filament parameters have on the profiles cannot be straightforwardly mimicked by choices of modifiers for scaling laws and parameters, especially while also ensuring consistency with the confined plasma modelled by MIRA. Instead, building surrogates based on Reduced-Order Models (ROMs) is proposed.

Fig. 3 shows a diagram of the strategy to build surrogates using ROMs, exemplified with values specific to building TOKES surrogates. The TOKES collection is split into two libraries (*upper left corner*): of "training" cases (24, for ROM construction) and of "withheld" cases (8, for surrogate validation), to ensure ~30% of the simulation results are reserved for validation. Withheld cases are cherry-picked to not include any extreme value of the T and V sets, to promote interpolative models

[1] The plasma temperature refers to both electrons and ions, as expected for the plasma in the birthplace of filaments.

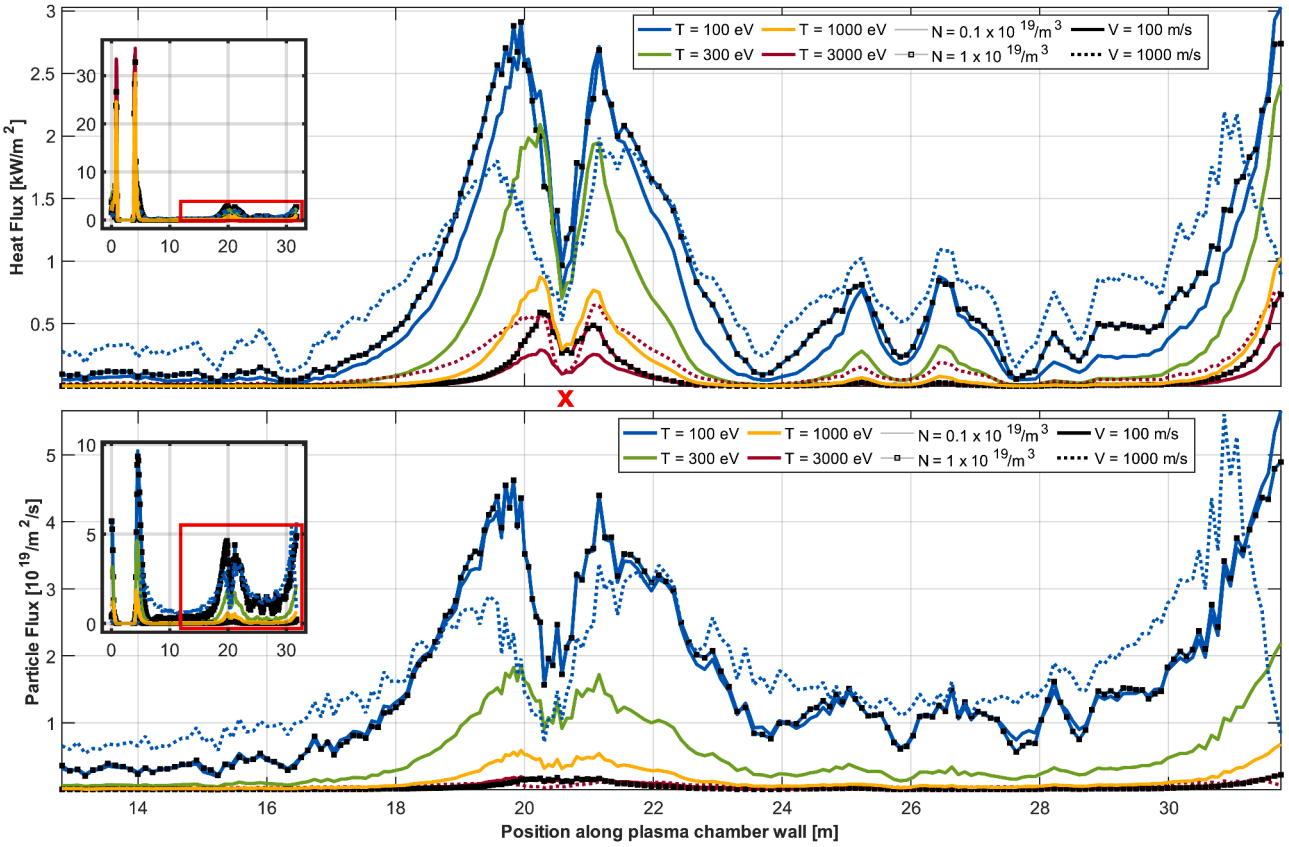


Fig. 2. Heat (*top*) and particle (*bottom*) flux profiles that reach the FW, for a subset of cases simulated by TOKES. Full profiles are seen miniaturized (*left*), while partial ranges of the same profiles (*red box*) are magnified to exemplify the impact of initial filament simulation parameters (*center*). Variations of the temperature ***T*** (*line color*), density ***N*** (*marker vs. no marker*) and ejection speed ***V*** (*continuous vs. dotted line*) alter, *e.g.*, the fraction of the profile deposited on the wall section close to the upper X-point; the letter label between plots (*red x*) indicates the same point of interest along the reactor chamber wall as in Fig. 1.

instead of extrapolative. Two datasets (for heat and particle profiles) are built from the training library, ordered into a matrix of "measurements" X and pre-processed into *standardized* data (mean-subtracted and with variance set to 1). Isomorphism parameters are stored for later inversion back to non-standardized space. Optionally, datasets can also be non-linearly transformed to improve the regression procedure tackled later; this is discussed in more detail in Section 4.

The PCA is subsequently applied; the technique expresses datasets of large dimensionality in a basis of singular vectors (components). These "modes" of data are ensured by the Singular Value Decomposition (SVD) algorithm to be orthogonal and have maximal correlation with the original datasets, ranked by their associated singular values. Linear combinations of all modes reconstruct the original results and define a Full Order Model (FOM). The amount of data variance reproduced by each mode is classified by its singular value, and data reconstruction with only the most meaningful modes defines a ROM instead. The full number of modes amounts up to the number of cases in the dataset [15]. Successful generalization of trends in the data with modes usually requires more cases than the dimensionality of a single case, but for highly structured data (*e.g.* human faces in classification algorithms, or far-SOL flux profiles that are strongly determined by the magnetic field geometry) this may be accomplished with considerably smaller datasets. This possibility must be assessed with ratios between singular values, which identify the number of modes required to represent a meaningful portion of the data variance (*e.g.* >90%) [19].

The resulting truncated set of modes (*i.e.* the ROM) is then used to build multivariate regression models against the filament parameters used to run each case, as shown in Fig. 3 (*green box*), similarly to a Principal Components Regression (PCR) but using the filament parameters directly as explanatory variables [20]. The use of (i) a ROM instead of a FOM decreases the chance of overfitting, *i.e.* increased model complexity that worsens its prediction capabilities, as does [19]:

(ii) a *k-fold cross-validation* procedure, by building k regression models out of random subsets of training cases (*upper right corner*); subset size (fold fraction) is usually about 70% of the training library. The average of each regression coefficient across all models defines the final coefficients for the surrogate model (*lower right corner*).

(iii) comparing models using Information (loss) Criteria, such as the Kullback-Leibler Divergence (KLD):

$$KLD = \int d_s(x,\beta,\mu)dx \tag{1}$$

$$d_s(x,\beta,\mu) = f(x,\beta)\cdot ln\left(\frac{f(x,\beta)}{g(x,\mu)}\right), \tag{2}$$

which measures the distance between the true profile $f(\cdots)$ and the predictor model $g(\cdots)$, or the loss of information ("surprise"), by integrating the Surprise Distribution d_s in the data space of X, given parameter sets β and μ. The quality of the surrogate is evaluated by comparing its predictions for profiles of each case (*i.e. each* (T,N,V) combination) with the original ones (*lower left corner*).

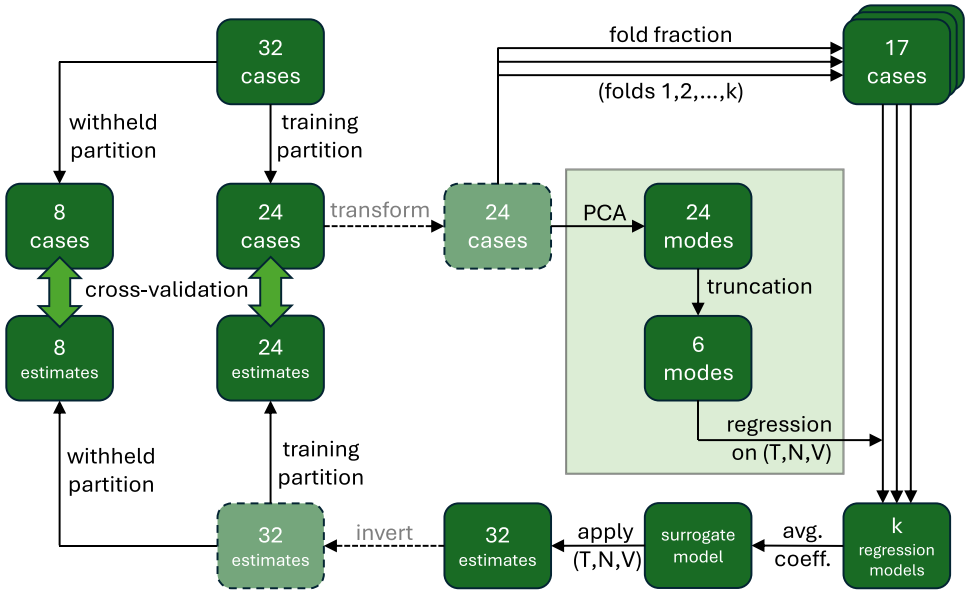


Fig. 3. Diagram of the strategy to build and validate each surrogate model. In this context, the term "partition" refers to a cherry-picked subset, while "fraction" refers to a random subset. Data transformation (*dashed lines/boxes*) are optional steps and have only been employed in the follow-up study discussed in Section 4.

3. Data-driven surrogates for TOKES

3.1. PCA of FW heat & particle flux profiles

Fig. 4 shows the first 6 PCA modes before inversion of standardization (*i.e.* nominal values not measured in $\vec{Q}$ and $\vec{G}$ units). Modes are plotted after projection in measurement space (*i.e. versus* the position along plasma chamber wall) to identify points of interest more easily. For example, as expected, peaks close to the Divertor and upper X-point are particularly visible on the first 2 modes, the ones that represent most of the variance captured by the PCA.

Fig. 5 shows that these first 6 modes explain more than 90% of the data variances, which corroborates the hypothesis that a limited number of TOKES simulations can be used to build regression-based surrogates due to the strong impact of the magnetic field on the profiles. Filaments propagate following magnetic field lines and necessarily connect with solid surfaces (leading to fast deposition of particles and heat) where these lines are designed to meet the FW; this results in fluxes larger at certain wall positions, such as near the upper and lower X-points.

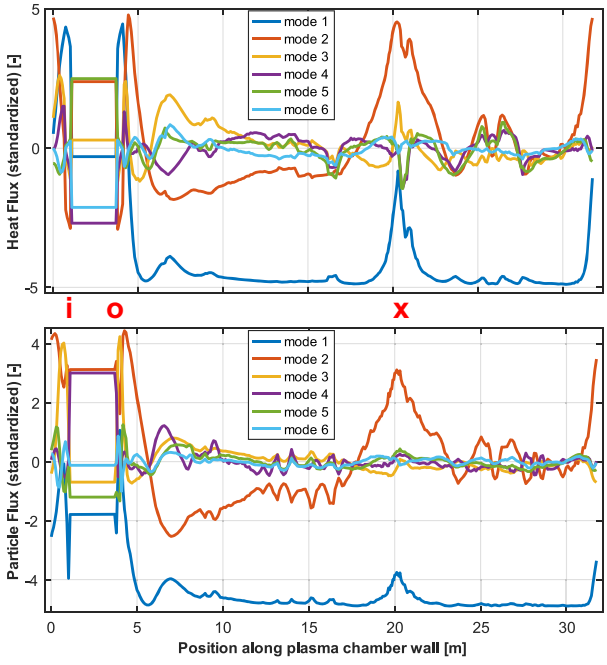


Fig. 4. Visualization of PCA modes in measurement space for standardized (*i.e.* dimensionless) data of heat and particle flux profiles. Letter labels (*red*) indicate the same points of interest along the reactor chamber wall as in Fig. 1.

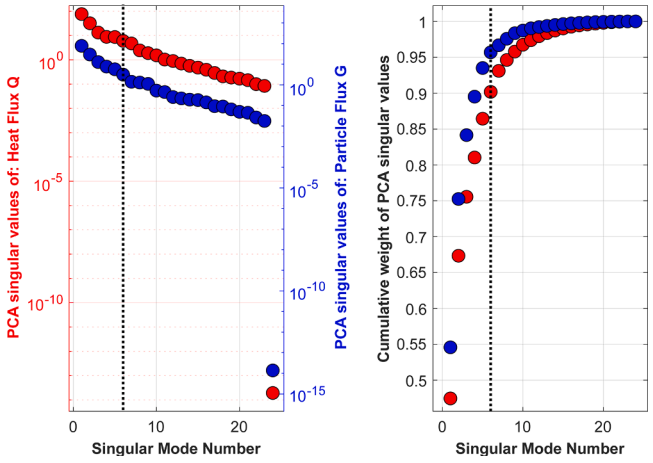


Fig. 5. Visualization of relevance of all PCA modes for heat fluxes (*red*) and particle fluxes (*blue*): scree plot for the singular values associated to each mode (*left*); ratio between cumulative and total sums of ranked singular values (*right*). For the 6 largest singular values, the associated modes already reproduce >90% of the data variance present in both datasets and suggest truncation (*dotted lines*) is possible.

3.2. Prediction of profiles by surrogate models

Only the first 6 PCA modes were used to build surrogate models through k-fold regression (*vide* Fig. 2). The number of folds k was determined by studying the change in the averaged regression coefficients for varying values of k between 2 and 2000. It was chosen as

the minimum for which deviation between the cumulative and the (3-elements) moving average for all averaged regression coefficients was below 10%; it was consistently computed to be between 20 and 40 folds.[2]

Fig. 6 shows a comparison between absolute values of KLD evaluations for the predictions of PCA and surrogate models for each case simulated by TOKES. To characterize each curve by a single numerical evaluation, the figure also displays geometrically averaged values of the KLD evaluations for all cases ("geo-mean", given the logarithmic spread). The curves and geo-means emphasize the following:

- the FOM reconstruction of the original TOKES profiles is associated with a surprise value that is computationally indistinguishable from zero.
- the ROM reconstruction of the original TOKES profiles is associated with surprise values between orders of magnitude 10^{-5} and 10^{-2}, and its geo-mean of 10^{-3}can be considered a minimum baseline for model comparisons, probably given by an unavoidable loss of information due to PCA truncation.
- the predictions for the training library (cases 1–24) computed by the surrogate models incur, on average, in surprise values of one order of magnitude higher than the reconstructions by the ROMs and coincide with most surprises for the withheld library (cases 25–32).

Fig. 7 can be used to further assess the capability of the surrogate models in predicting profiles. It shows two examples of predictions when compared to the original profiles computed by TOKES: one from the training library (case 2) and one from the withheld library (case 27). Of note:

- most of the necessary profile shapes are well-captured by the surrogates, with just some notable deviation in the local maxima close to the Divertor.
- as expected from the original profiles, predicted values inside the Divertor region are null.

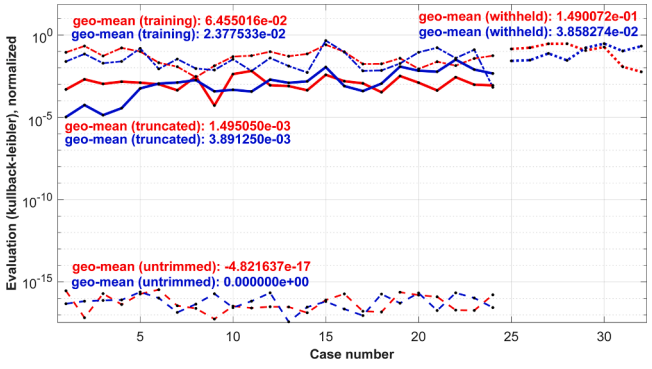


Fig. 6. Model evaluations using the KLD, for heat flux (*red*) and particle flux (*blue*) profiles, case-by-case, alongside geometric-averages for all cases of each library and model. By increasing order of magnitude of surprise: FOM reconstruction of TOKES profiles (*dashed lines*, "untrimmed"); ROM reconstruction of TOKES profiles (*continuous lines*, "truncated"); averaged-regression model prediction of training cases (*dash-dotted lines*, "training"); averaged-regression model prediction of withheld cases (*dotted lines*, "withheld").

[2] The value k must not be strictly optimized, just big enough to allow for a relatively stable set of averaged regression coefficients. Some lower values of k might even produce better evaluation results, but it is statistically likely these represent models that are good at training-data reconstruction, but bad at predicting unknown profiles, and thus were avoided.

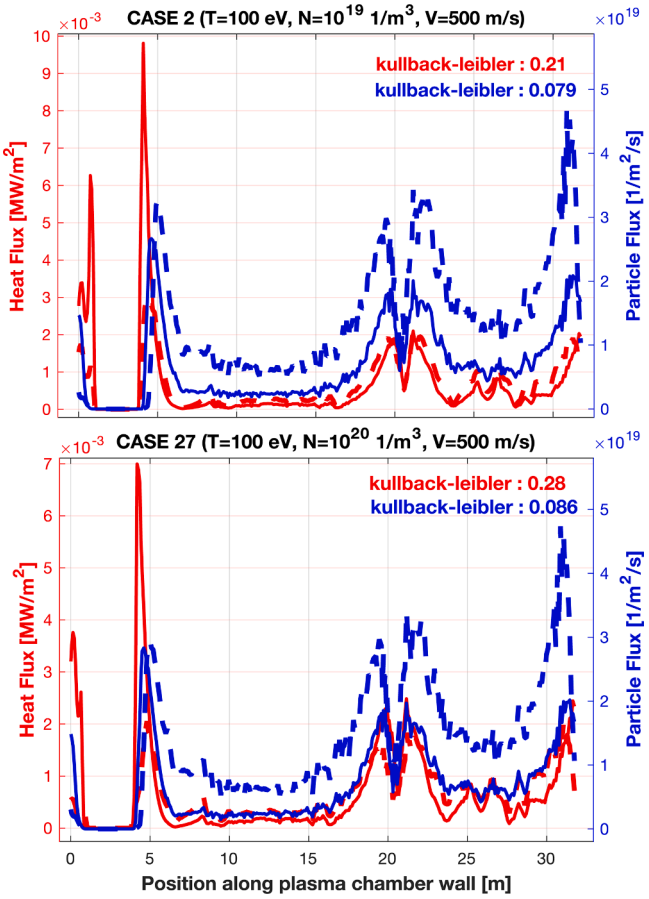


Fig. 7. Comparison between original profiles simulated by TOKES (*continuous lines*) and predicted profiles (*dashed lines*), for the heat (*red*) and particle (*blue*) fluxes, in two representative cases: a training case (2) and a withheld case (27). Evaluations with the KLD of the predicted profiles computed with the surrogate models are displayed in the corners.

- surrogate responses are similar for similar (T, N, V) combinations, even for combinations not used for model training, which is mathematically represented by the similar KLD evaluations between training and withheld cases.

These, alongside the evaluations in Fig. 6, imply that the surrogate models presented so far can successfully reproduce the profiles simulated by TOKES, albeit with the incursion of non-negligible error; the optimization procedure presented in Section 4 is proposed to potentially reduce this error.

4. Optimization of TOKES surrogates

Surrogate models based on multivariate regression might not be able to capture all data trends, such as non-linear dependencies. However, Koopman Operator theory suggests that it is possible to find an appropriate choice of dataset transformation prior to the application of the PCA that linearly represents the system, depicted as the optional step in Fig. 3 (*dashed boxes*) [19]. Heat and particle fluxes due to filaments might correlate, for example, with the electron collisionality in the far-SOL, which scales with $N/(T_e)^2$ and whose low values is known to suppress filaments by shearing [10].

With this assumption, the elements $X_{p,c}$ of the training matrix were transformed:

$$\widetilde{X}_{p,c} = X_{p,c} \cdot (T_c)^t \cdot (N_c)^n \cdot (V_c)^v, \tag{3}$$

for all indexes representing positions p along the chamber wall and cases c of the training library. For varying combinations(t,n,v) of these exponents, the exponent-space of $\varepsilon \in [-10,10]$ was scanned for each exponent $\varepsilon = t,n,v$, with the goal of determining for which combination the geo-mean evaluation of each library is minimized (*i.e.* the surrogates that most correctly predict all TOKES profiles). For this scan, a Modified Divergence MD was defined:

$$MD = \left| \int d_m(x,\beta,\mu)dx \right| \tag{4}$$

$$d_m(x,\beta,\mu) = \begin{cases} f(x,\beta), & \forall\ (g=0) \\ g(x,\beta), & \forall\ (f=0) \wedge (g \neq 0) \\ d_s(x,\beta,\mu), & otherwise, \end{cases} \tag{5}$$

since the conventional KLD Surprise Distribution d_s (*vide* Eq. (2)) is not defined where surrogates predict regions with zero flux.[3] The Modified Distribution d_m (*vide* Eq. 5) ensures that discrepancies between the original and the predicted profiles are more strongly penalized, wherever exclusively one of them displays non-trivial null-flux regions.

Fig. 8 shows 3D plots of some selected results of evaluations using the MD during these exponent scans; as an example, evaluation geo-means are displayed only for the training cases. In both graphs, a point can be seen where the geo-mean is evidently minimized, found at the intersection where the geo-mean is lowest for each single-parameter scan ("best" mid-planes). This strongly suggests an optimal transformation exists.

Fig. 9 shows the search for lowest evaluation geo-means, when plotted against the variation of each single exponent. Each curve fixes two exponents and varies the last one within the exponent-space of ε, with a step of $\Delta\varepsilon = 0.02$ within $[-1,1]$ and with a step of $\Delta\varepsilon = 1$ for the rest. All curves identify a geo-mean minimum along its associated single-exponent scan, and that minimum determines the exponent value that defines the associated "best geo-mean" mid-plane. These minima identify the data transformations that best linearize the datasets in respect to (T,N,V) for each library, with an uncertainty given by the exponent step size $\Delta\varepsilon$ around each minimum, and are listed in Table 1. Results from training and withheld libraries were kept separate and optimal data transformations were estimated as their weighted average using the number of cases in each library as weights, again to promote interpolative models. Such weighted averages were used to build optimized surrogates.

Fig. 10 shows predictions for the same training and withheld cases given as examples in Section 3.2 (2 and 27, respectively), but produced by the optimized surrogates. Notably, the error between prediction and original profiles is considerably reduced, except for the discrepancies

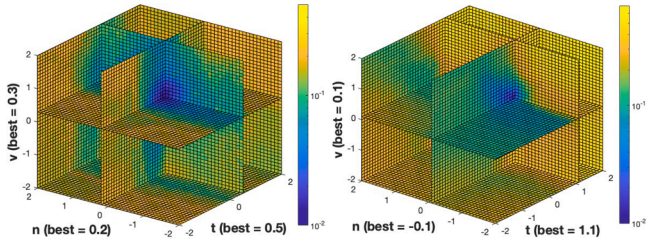


Fig. 8. Visualization of the distribution of (MD) evaluations geo-mean for the prediction of training cases for the heat flux (*left*) and particle flux (*right*) profiles in exponent-space. Results, displayed for an exponent scan in the $\boldsymbol{\varepsilon} \in [-2,2]$ range with a step of $\Delta\boldsymbol{\varepsilon} = 0.1$, show a point (intersection between "best geo-mean" mid-planes) in which the geo-mean is globally minimized.

[3] For $f(x,\cdots) = 0$, one can assume $d_s(x,\cdots) = 0$, but the same is not true for $g(x,\cdots) = 0$. Alternative failed attempts included variants of the KLD (*e.g.* Jeffreys, Jensen-Shannon) and treating the profiles (*e.g.* substitution of predictions by their moving-averaged versions before evaluation).

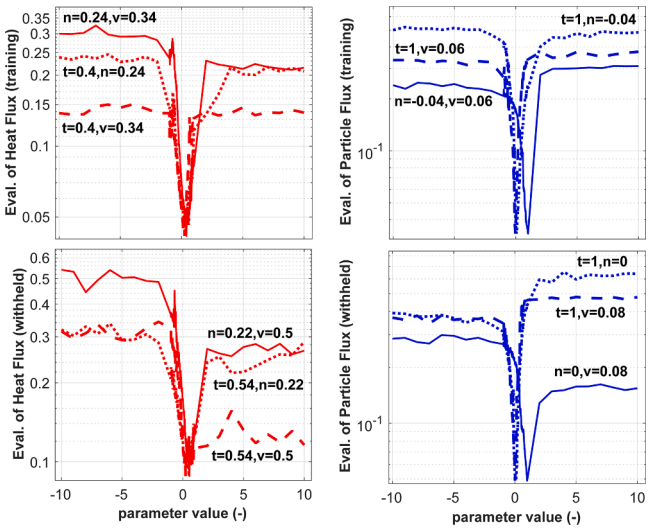


Fig. 9. Visualization of the geo-mean variation within each single-exponent scan, when keeping the other two exponents fixed at the values where they achieve minimum geo-mean during their own scans. Evaluations plotted separately for training (*top*) and withheld (*bottom*) cases, for both heat (*left*) and particle (*right*) flux profile predictions.

Table 1
Exponent values associated with the global minimum of the geo-mean (MD) evaluation scan for each flux profile (heat & particle) and library (training & withheld).

flux	ε	exponent value (±0.02) [-]		
		training (x24)	withheld (x8)	average (case-weighted)
$\vec{Q}$	t	0.40	0.54	0.43
	n	0.24	0.22	0.23
	v	0.34	0.50	0.38
$\vec{G}$	t	1.00	1.00	1.00
	n	− 0.04	0.00	− 0.03
	v	0.06	0.08	0.06

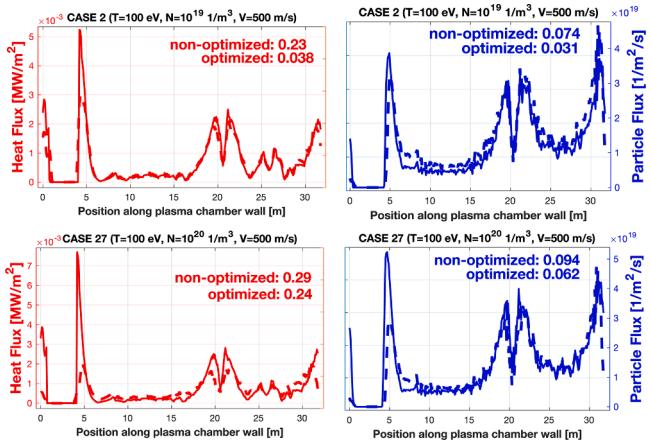


Fig. 10. Comparison between original profiles simulated by TOKES (*continuous lines*) and profiles predicted by optimized surrogates (*dashed lines*), for the heat (*red*) and particle (*blue*) fluxes, for (training) case 2 (*top*) and (withheld) case 27 (*bottom*). Evaluations performed with the MD are displayed in the corners ("optimized"), alongside the MD evaluations for the base surrogates ("non-optimized").

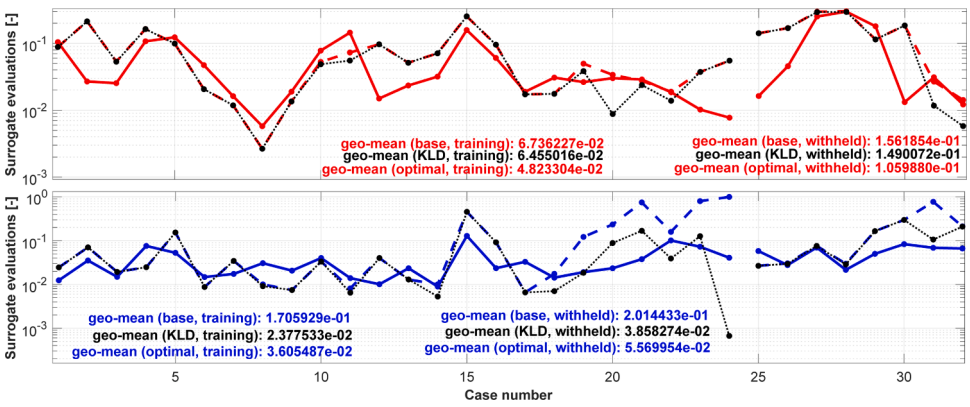


Fig. 11. Model MD evaluations for each case and their geo-means, for heat fluxes (*top, red*) and particle fluxes (*bottom, blue*) profiles, for two different surrogates: base (*dashed lines*), built with non-transformed data; and optimized (*continuous lines*) built with data transformed with exponents in Table 1. For comparison, base surrogate evaluations and geo-means are also provided for KLD evaluations in both graphs (*dotted lines, black*).

close to the Divertor region. This indicates that data transformation is largely successful in obtaining a better set of surrogate models.

Fig. 11 shows the MD evaluations of predictions produced with the optimized surrogates (*continuous lines*). For comparison, the evaluations for the non-optimized surrogates are also displayed, computed with the MD (*dashed lines*) and the KLD (*dotted lines*). Comparison between the latter ones shows that KLD and MD evaluations coincide in most points; when not, the MD is always larger because it adds significant surprise wherever predictions either do not match null-regions in the profiles, or estimate zero for non-null regions. Furthermore, MD evaluations for optimized and non-optimized surrogates are similar for most cases, but are considerably better for others (*e.g.* 2, 13, 24, 25, 30, 31); the overall improvement of the capabilities of optimized surrogates is represented by smaller geo-means.

5. Discussions & outlook

Surrogates for the simplified turbulent transport model TOKES were developed using multi-variate regression on the modes of a Principal Components Analysis (PCA), with the goal of enabling consistent systemic analyses in SCs that include filament far-SOL dynamics and heat/particle flux profiles, unfeasible with scaling laws. This was done by presenting a systematic methodology to build and optimize these surrogates, including a novel Modified Divergence (MD) expression used to evaluate their capabilities.

The use of regression was proposed possible, despite the limited number of different TOKES simulations, due to the strong structuring of the flux profiles caused by the magnetic field geometry. This hypothesis was verified by the relatively low number of PCA modes necessary to represent more than 90% of the data variance of the training library for both profiles, even for the large ranges of initial filament parameters used as simulation inputs, which validated the approach.

The optimization process relied on the assumption that the flux profiles could be linearized in relation to the initial filament parameters (T, N, V) by an appropriate choice of data transformation prior to the PCA. Scans over an exponent-space for powers of these parameters searched for the transformation that minimized model evaluations with the MD, which was developed to counteract limitations of conventional expressions. This revealed the exponents collected in Table 1 and implied the following dependencies:

$$\vec{P}\ T^t N^n V^v = \sum_i \overrightarrow{m_i}\ \left(c_{T,i}T + c_{N,i}N + c_{V,i}V\right), \tag{6}$$

with $\vec{P}$ representing either the heat flux profile $\vec{Q}$ or the particle flux profile $\vec{G}$, when $\overrightarrow{m_i}$ are the PCA profile modes associated with that flux, and $c_{\dots,i}$ are the likewise regression coefficients. Notably, these do not explicitly display the electron collisionality, which suggests that its correlation to filament behavior does not extend to their subsequent interaction with the walls. On the other hand, this result denotes the potential insights on unknown dependencies that the technique developed in this work may reveal when applied to results of physics-based simulations. These patterns may, in turn, be compared with modeling efforts to justify or challenge them.

Before optimization, predictions by surrogates incur in non-negligible error when compared to the original profiles. However, it should be reiterated that all TOKES profiles are normalized to a prescribed far-SOL power of 1 *MW*, *i.e.* their application requires re-normalization. Likewise, predicted profiles must also be renormalized by a far-SOL transport power computed by a confined plasma model, which implies a stronger importance of their shape over their nominal values.

After optimization, the error in predictions was significantly reduced throughout most of each profile, with prevailing discrepancies on the peaks centered around the Divertor region. Arguably, however, this deviation is not significant, since the surrogate models are aimed at SCs; tools notoriously able to cope with relatively big deviations when the impact is not meaningful on parametric and scoping analyses, or integrated design optimization. In qualitative terms, the absolute error incurred by these discrepancies in the predictions of far-SOL fluxes near the Divertor is likely considerably limited due to: (1) the narrow region in the chamber wall surface with which it is associated (and far from the equatorial line, which minimizes the effect of profile revolution); and (2) the extent of its contribution in comparison to near-SOL transport and radiation contributions.

In fact, current experimental prescriptions distribute the total transport power across the separatrix approximately equally between near-SOL and far-SOL regions [3], if not considerably biased towards the near-SOL [2]. Considering that the wall surface area of the Divertor makes up for less than 9% of the total reactor chamber wall area (geometry referring to PROCESS SC modeling [21,22]), an application of the Eich scaling law for the near-SOL (such as the one currently implemented in MIRA [14]) produces fluxes of at least one order of magnitude higher than those of the surrogate predictions, which effectively minimizes the impact of far-SOL deviations on total SOL loads.

In the framework of the MIRA SC simulating a fixed EU-DEMO reactor design (that is, for the fixed plasma equilibrium and wall geometry used to run TOKES), these optimized surrogates are expected to perform novel parametric analyses of the impact of filament properties to other power plant systems. This will enable the study of potentially relevant dependencies, such as between the design temperature of the FW and effusion fluxes during pump-down, given the (consistent) heat and particle load profiles. Initial filament temperatures and densities are planned to be derived from the confined plasma model in MIRA, while

ejection velocities must be pre-defined (substituting the decay length in scaling models as user parameter). As seen in Fig. 2, the latter can redistribute heat and particles fluxes along the chamber wall and an assessment of its systems-level impact is also desirable, which cannot be achieved with conventional scaling laws. Ideally, empirical reports for the partitioning of filaments characterization should be used to describe their population in the SOL; but while unavailable, extreme values based on literature measurements can be conservatively used.

Further improvement of the proposed methodology can be pursued with techniques like gradient-descent in the exponent-space scan, or SINDy to identify alternative transformations. Additionally, while a far-SOL model compatible with SC is not available, generalization of the surrogates to other magnetic field operational points can be investigated, by including plasma and coils currents as appended regression parameters.

CRediT authorship contribution statement

Tiago Pomella Lobo: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sergey Pestchanyi:** Writing – review & editing, Supervision, Software, Resources, Investigation, Data curation, Conceptualization. **Ivan Alessio Maione:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tiago Pomella Lobo reports financial support was provided by European Consortium for the Development of Fusion Energy. Tiago Pomella Lobo reports a relationship with European Consortium for the Development of Fusion Energy that includes: funding grants.

Data availability

Data will be made available on request.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement N° 101052200 — EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

References

[1] P.C. Stangeby, The Plasma Boundary of Magnetic Fusion Devices, Institute of Physics Pub, Bristol ; Philadelphia, 2000.
[2] R. Wenninger, R. Albanese, R. Ambrosino, F. Arbeiter, J. Aubert, C. Bachmann, L. Barbato, T. Barrett, M. Beckers, W. Biel, L. Boccaccini, D. Carralero, D. Coster, T. Eich, A. Fasoli, G. Federici, M. Firdaouss, J. Graves, J. Horacek, M. Kovari, S. Lanthaler, V. Loschiavo, C. Lowry, H. Lux, G. Maddaluno, F. Maviglia, R. Mitteau, R. Neu, D. Pfefferle, K. Schmid, M. Siccinio, B. Sieglin, C. Silva, A. Snicker, F. Subba, J. Varje, H. Zohm, The DEMO wall load challenge, Nucl. Fusion 57 (2017) 046002, https://doi.org/10.1088/1741-4326/aa4fb4.
[3] F. Maviglia, M. Siccinio, C. Bachmann, W. Biel, M. Cavedon, E. Fable, G. Federici, M. Firdaouss, J. Gerardin, V. Hauer, I. Ivanova-Stanik, F. Janky, R. Kembleton, F. Militello, F. Subba, S. Varoutis, C. Vorpahl, Impact of plasma-wall interaction and exhaust on the EU-DEMO design, Nucl. Mater. Energy 26 (2021) 100897, https://doi.org/10.1016/j.nme.2020.100897.
[4] L.V. Boccaccini, F. Arbeiter, P. Arena, J. Aubert, L. Bühler, I. Cristescu, A.D. Nevo, M. Eboli, L. Forest, C. Harrington, F. Hernandez, R. Knitter, H. Neuberger, D. Rapisarda, P. Sardain, G.A. Spagnuolo, M. Utili, L. Vala, A. Venturini, P. Vladimirov, G. Zhou, Status of maturation of critical technologies and systems design: breeding blanket, Fusion Eng. Des. 179 (2022) 113116, https://doi.org/10.1016/j.fusengdes.2022.113116.
[5] R. Arredondo, K. Schmid, F. Subba, G.A. Spagnuolo, Preliminary estimates of tritium permeation and retention in the first wall of DEMO due to ion bombardment, Nucl. Mater. Energy 28 (2021) 101039, https://doi.org/10.1016/j.nme.2021.101039.
[6] T. Eich, B. Sieglin, A. Scarabosio, W. Fundamenski, R.J. Goldston, A. Herrmann, A. S.D.E.X. Upgrade Team, Inter-ELM power decay length for JET and ASDEX upgrade: measurement and comparison with heuristic drift-based model, Phys. Rev. Lett. 107 (2011) 215001, https://doi.org/10.1103/PhysRevLett.107.215001.
[7] T. Eich, A.W. Leonard, R.A. Pitts, W. Fundamenski, R.J. Goldston, T.K. Gray, A. Herrmann, A. Kirk, A. Kallenbach, O. Kardaun, A.S. Kukushkin, B. LaBombard, R. Maingi, M.A. Makowski, A. Scarabosio, B. Sieglin, J. Terry, A. Thornton, A.S.D. E.X. Upgrade Team, JET EFDA Contributors, Scaling of the tokamak near the scrape-off layer H-mode power width and implications for ITER, Nucl. Fusion 53 (2013) 093031, https://doi.org/10.1088/0029-5515/53/9/093031.
[8] F. Maviglia, G. Federici, R. Wenninger, R. Albanese, R. Ambrosino, C. Bachmann, L. Barbato, F. Cismondi, M. Firdaouss, V.P. Loschiavo, C. Lowry, Effect of engineering constraints on charged particle wall heat loads in DEMO, Fusion Eng. Des. 124 (2017) 385–390, https://doi.org/10.1016/j.fusengdes.2017.02.077.
[9] M. Giacomin, A. Stagni, P. Ricci, J.A. Boedo, J. Horacek, H. Reimerdes, C.K. Tsui, Theory-based scaling laws of near and far scrape-off layer widths in single-null L-mode discharges, Nucl. Fusion 61 (2021) 076002, https://doi.org/10.1088/1741-4326/abf8f6.
[10] W. Zholobenko, J. Pfennig, A. Stegmeir, T. Body, P. Ulbl, F. Jenko, Filamentary transport in global edge-SOL simulations of ASDEX Upgrade, Nucl. Mater. Energy 34 (2023) 101351, https://doi.org/10.1016/j.nme.2022.101351.
[11] S. Peschanyy, M. Lehnen, A. Huber, I. Landman, Verification of TOKES simulations against the MGI experiments in JET, Fusion Eng. Des. 87 (2012) 1195–1200, https://doi.org/10.1016/j.fusengdes.2012.02.107.
[12] S. Peschanyy, Simulation of Heat Flux to the DEMO First Wall Due to Filamentary Transport in the Far SOL, IEEE Trans. Plasma Sci. 46 (2018) 1393–1397, https://doi.org/10.1109/TPS.2017.2781191.
[13] F. Franza, L.V. Boccaccini, E. Fable, I. Landman, I.A. Maione, S. Petschanyi, R. Stieglitz, H. Zohm, MIRA: a multi-physics approach to designing a fusion power plant, Nucl. Fusion 62 (2022) 076042, https://doi.org/10.1088/1741-4326/ac6433.
[14] F. Franza, Development and Validation of a Computational Tool For Fusion Reactors' System Analysis, Karlsruher Institut für Technologie, 2019.
[15] I.T. Jolliffe, Principal Component Analysis, 2nd ed, Springer, New York, 2002.
[16] R. Andrews, J. Diederich, A.B. Tickle, Survey and critique of techniques for extracting rules from trained artificial neural networks, Knowl. Based. Syst. 8 (1995) 373–389, https://doi.org/10.1016/0950-7051(96)81920-4.
[17] G. Marcus, Deep learning: a critical appraisal, (2018). https://doi.org/10.48550/ARXIV.1801.00631.
[18] A.M. Zador, A critique of pure learning and what artificial neural networks can learn from animal brains, Nat. Commun. 10 (2019) 3770, https://doi.org/10.1038/s41467-019-11786-6.
[19] S.L. Brunton, J.N. Kutz, Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control, 1st ed., Cambridge University Press, 2019 https://doi.org/10.1017/9781108380690.
[20] A.S. Hadi, R.F. Ling, Some Cautionary Notes on the Use of Principal Components Regression, Am. Stat. 52 (1998) 15–19, https://doi.org/10.1080/00031305.1998.10480530.
[21] M. Kovari, R. Kemp, H. Lux, P. Knight, J. Morris, D.J. Ward, PROCESS": a systems code for fusion power plants - Part 1: physics, Fusion Eng. Des. 89 (2014) 3054–3069, https://doi.org/10.1016/j.fusengdes.2014.09.018.
[22] M. Kovari, F. Fox, C. Harrington, R. Kembleton, P. Knight, H. Lux, J. Morris, PROCESS": a systems code for fusion power plants – Part 2: engineering, Fusion Eng. Des. 104 (2016) 9–20, https://doi.org/10.1016/j.fusengdes.2016.01.007.