

Data-driven model for divertor plasma detachment prediction

Ben Zhu^{1,†}, Menglong Zhao¹, Harsh Bhatia¹, Xue-qiao Xu¹,
Peer-Timo Bremer¹, William Meyer¹, Nami Li¹ and Thomas Rognlien¹

¹Lawrence Livermore National Laboratory, Livermore, CA 94550, USA

(Received 9 June 2022; revised 2 September 2022; accepted 2 September 2022)

We present a fast and accurate data-driven surrogate model for divertor plasma detachment prediction leveraging the latent feature space concept in machine learning research. Our approach involves constructing and training two neural networks: an autoencoder that finds a proper latent space representation (LSR) of plasma state by compressing the multi-modal diagnostic measurements and a forward model using multi-layer perception (MLP) that projects a set of plasma control parameters to its corresponding LSR. By combining the forward model and the decoder network from autoencoder, this new data-driven surrogate model is able to predict a consistent set of diagnostic measurements based on a few plasma control parameters. In order to ensure that the crucial detachment physics is correctly captured, highly efficient 1D UEDGE model is used to generate training and validation data in this study. The benchmark between the data-driven surrogate model and UEDGE simulations shows that our surrogate model is capable of providing accurate detachment prediction (usually within a few per cent relative error margin) but with at least four orders of magnitude speed-up, indicating that performance-wise, it has the potential to facilitate integrated tokamak design and plasma control. Comparing with the widely used two-point model and/or two-point model formatting, the new data-driven model features additional detachment front prediction and can be easily extended to incorporate richer physics. This study demonstrates that the complicated divertor and scrape-off-layer plasma state has a low-dimensional representation in latent space. Understanding plasma dynamics in latent space and utilising this knowledge could open a new path for plasma control in magnetic fusion energy research.

Key words: fusion plasma, plasma confinement

1. Introduction

One of the critical concerns for a magnetic fusion power plant is the intense plasma heat flux q_{\parallel} that flows very rapidly along magnetic field lines to localised material divertor plates. For unmitigated conditions, this heat flux can exceed 10 GW m^{-2} (Kuang *et al.* 2020) based on extrapolations from current experimental databases and challenges even

† Email address for correspondence: zhu12@llnl.gov

the most highly developed modern materials with thermo-mechanical engineering limits at around 10 MW m^{-2} . So far, the most successful method of reducing this peak heat flux in tokamaks is achieved by *detachment*; that is, the use of neutral gas to buffer the incoming plasma heat flux and radiate much of the incoming power over a much wider area of the surrounding walls as the recombined neutrals and the radiation photons freely stream across the magnetic field rather than being focused to the plate as for the plasma. In other words, upstream high-temperature plasma is extinguished by high-density neutral gas before it strikes the divertor plates; hence, it is *detached*.

Because the detachment phenomenon involves nonlinear coupled atomic, molecular and plasma physics, it is challenging to obtain accurate and fast detachment predictions (i.e. predict divertor plasma quantities such as density and temperature for prescribed upstream plasma condition). On the one hand, theory-based models with simplifications such as the semi-analytical two-point model formatting (2PMF) (Stangeby 2018) derives analytical formula to calculate divertor plasma density and temperature. These over-simplified zero-dimensional models are fast and valuable in terms of providing physics insights, but the corresponding prediction is somewhat crude and relies on *ad-hoc* or data-fitted momentum and energy loss fraction coefficients. On the other hand, sophistic tokamak edge transport codes, such as two-dimensional (2D) UEDGE (Rognien *et al.* 1994) and SOLPS-ITER (Wiesen *et al.* 2015) that incorporate a higher level of complexity and physical effects could give fairly accurate predictions but require a few days to a few months to finish one run. Of course, with a large enough neutral gas and/or impurity injection, detachment almost always occurs; however, ‘too much’ detachment can lead to disruption, a rapid termination of plasma confinement, which could also damage the machine. Thus, predicting, establishing and maintaining a proper degree of plasma detachment, along with a stable and properly positioned ‘detachment front’, is a non-trivial requirement. To make this challenge ever greater, current detachment control in existing tokamaks is accomplished only with the aid of *in situ* diagnostics such as infrared camera (Maurizio *et al.* 2017) or Langmuir probe (Eldon *et al.* 2022) that would not survive in a reactor environment despite that the success of future magnetic fusion reactors, such as ITER, rely on detachment to mitigate excessive heat load on divertor targets. Therefore, attaining a reliable and fast detachment prediction model and control algorithm is imperative to magnetic fusion energy research.

To address this challenge, we propose to build a detachment prediction model with the data-driven machine learning (ML) approach. For the last few years, ML methods have gained a lot of interest in magnetised plasma research, as summarised in the latest review paper by Anirudh *et al.* (2022). Various neural networks have been explored and applied to assorted applications of experimental and computational fusion plasma study, e.g. to accelerate data processing (van den Berg *et al.* 2018), to interpret critical plasma parameter directly from diagnostics (Samuell *et al.* 2021), to aid plasma shaping via magnetic field control (Degraeve *et al.* 2022), to forecast disruption (Kates-Harbeck, Svyatkovskiy & Tang 2019; Montes *et al.* 2019; Rea *et al.* 2020), to speed-up the nonlinear Fokker–Planck–Landau collision operator in particle simulations (Miller *et al.* 2021), to apply closures in fluid models (Ma *et al.* 2020; Maulik *et al.* 2020; Wang *et al.* 2020) and to discover governing physics-based partial observations (Mathews *et al.* 2021). The growing applications of ML techniques on plasma physics indicates that data-driven science could provide an alternative approach to scientific challenges.

In this paper, we report a fast yet accurate model-based data-driven detachment prediction model. The rest of the paper is organised as follows. The methodology of our approach is discussed in § 2. Section 3 describes the one-dimensional (1D) UEDGE model used in the study and outlines the data-generation process. Section 4 presents the

architecture of neural networks and hyper-parameters used in this study. The results of our trained models are disseminated in § 5. Section 6 compares 1D UEDGE simulations and our predictive model with the basic two-point model (2PM) and 2PMF. Application of our model, its limitations and future work are briefly discussed in § 7. Finally, § 8 summarises the paper.

2. Methodology

Steady-state plasma states are described in two different ways: (1) by control parameters x or (2) by diagnostic measurements y . For example, if we neglect any bifurcation or hysteresis phenomena for now, a tokamak plasma is determined by a complete set of engineering and discharge parameters or can be characterised with of diagnosed plasma quantities in detail. Analogously, in numerical simulations, a set of model inputs with proper boundary conditions yields a plasma state that can also be diagnosed synthetically. A large portion of magnetised plasma research, in fact, has been aimed at discovering and exploiting the relation between these two descriptions, i.e. predicting laboratory and fusion plasma behaviour under certain discharge parameters or finding an optimal control setting for a desired plasma state. In most situations, this is not a trivial task because of the strongly nonlinear nature of magnetised plasmas. On the one hand, sophisticated numerical models are developed to better understand the physics which governs the complex plasma system, and hopefully to achieve true first-principle-based predictive capability. On the other hand, empirical and simplified models are proposed to provide tools for daily machine operations, experiment planning and future device design. Often there is a gap in between them: empirical and simplified models are fast but not necessarily accurate, whereas sophisticated numerical models can provide more reliable predictions but too slow to be used for practical application purposes. A fast yet accurate surrogate model can bridge this gap. In this paper, we develop such a surrogate model for divertor plasma detachment prediction with a data-driven approach.

Unlike most data-driven surrogate models that directly connect two states x and y , we take the indirect approach proposed by (Anirudh *et al.* 2020). As shown in figure 1, in addition to two conventional descriptions of plasma, here we propose a third description, that in a latent space, termed as latent space representation (LSR). Latent space, sometimes also referred to as feature space, is a widely used concept in ML research in which items with similar features are positioned closely. We first find the LSR, z , of plasma by compressing diagnostics (e.g. synthetic diagnostics mimicking physical diagnostics, such as Langmuir probe, Thomson scattering and bolometer/radiation measurement) through an autoencoder (AE) (described in § 4.1); then forward and inverse models can be trained to relate control parameters x (engineering or numerical model input parameters such as heating power and gas puffing rate/upstream density) and z . Hence, instead of constructing neural networks $x \rightarrow z$ or $z \rightarrow x$, our forward and inverse predictions are $x \rightarrow z \rightarrow y$ and $y \rightarrow z \rightarrow x$. Compared with the direct approach, the study by Anirudh *et al.* (2020) showed that the indirect approach has several advantages, including improved predictive performance and greater data efficiency. In this paper, we focus on forward prediction (i.e. $x \rightarrow z \rightarrow y$), whereas the inverse prediction (optimisation problem) is subject to future work.

3. Physics model and data

In this section, we briefly describe the tokamak edge model used to simulate divertor detachment and outline the data-generation process.

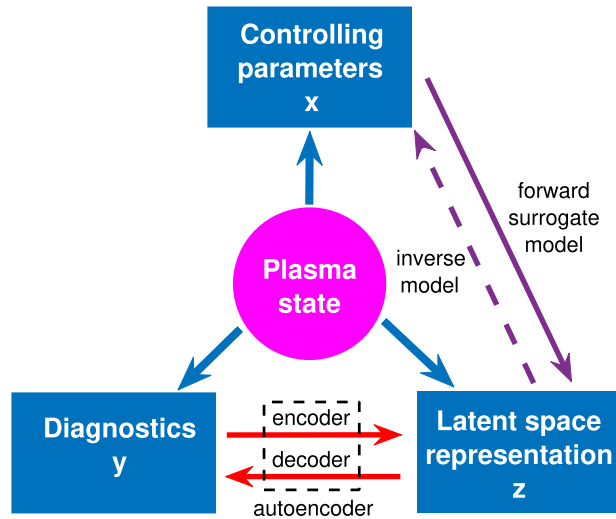


FIGURE 1. Three descriptions of plasma state.

3.1. 1D UEDGE model

UEDGE (Rognien *et al.* 1994; Rognien & Rensink 2002) is a finite-volume simulation code to capture plasma and neutral transport. UEDGE has been used extensively to model the edge region of present-day and future tokamak devices for almost three decades. It has been utilised to study the divertor detachment physics on many machines, such as Alcator C-Mod (Wising *et al.* 1997), DIII-D (Porter *et al.* 1996), ITER (Wising *et al.* 1996) and SPARC (Ballinger *et al.* 2021). UEDGE solves the multi-species fluid transport model for plasma and neutral gas, including important atomic physics, such as line radiation from ionisation, excitation and recombination processes in realistic tokamak equilibria. With the fully implicit Newton–Krylov scheme and adaptive stepping to advance all model equations in time, UEDGE is able to quickly evolve the system to a steady state.

In this proof-of-principle study, highly flexible and efficient 1D UEDGE code is used to quickly generate a large training data set that covers a wide range of parameter space beyond normal tokamak operation. Unlike the commonly used 2D UEDGE model, which simulates the entire tokamak periphery, including closed flux surface region, scrape-off-layer (SOL) and private flux surface region, the 1D UEDGE in this study models only one flux tube in the low-field side SOL, starting from the outer mid-plane to the outer divertor target plate, as depicted in figure 2(a). Nevertheless, 1D UEDGE code solves the same set of equations as its 2D counterpart except radial transport and drifts. Plasmas are allowed to transport in the parallel direction that is along the magnetic field whereas neutrals transport poloidally along the flux tube. As a result, the essence of plasma detachment physics is retained despite the one-flux-tube assumption. In fact, simplified 1D geometry disentangles many secondary effects and is sometimes preferred to better quantify the role of different physics processes played in the detachment (Dudson *et al.* 2019). In the 1D UEDGE set-up, spatial variation of the magnetic field \mathbf{B} is retained, i.e. the flux expansion along a flux tube in the SOL is considered in the 1D UEDGE mesh. Power, or the energy influx, across the core boundary is injected at the outer midplane (upstream), and is equally partitioned between electrons and ions. Plasma particles are also assumed to be fueled at upstream. As plasma moves downside the flux tube, it recombines

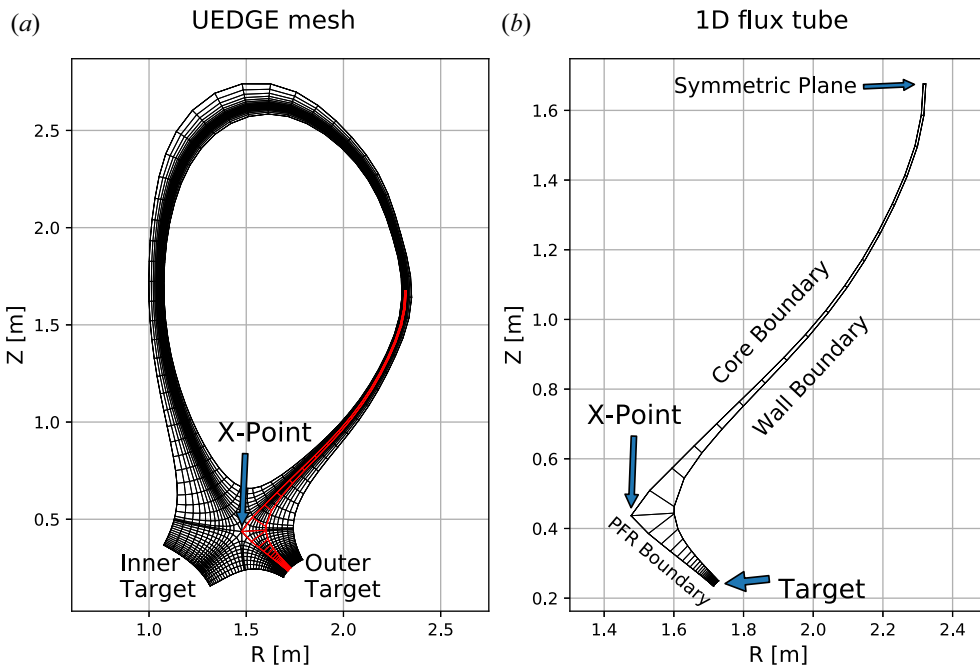


FIGURE 2. (a) 2D (black) versus 1D (red) UEDGE simulation meshes and (b) sketch of 1D UEDGE simulation set-up.

to neutrals which eventually are recycled at the divertor target plate. In this study, the particle and momentum recycling coefficients are set to be 0.99.

Figure 3 shows the typical attached (in blue) and detached (in orange) plasma in 1D UEDGE simulations. Compared with attached plasma, detached plasma normally has a lower electron temperature T_e , a higher neutral gas density n_g at the divertor target and a distinct detachment front, or radiation peak away from the divertor target. Not surprisingly, 1D UEDGE code is able to qualitatively reproduce the target ion saturation current density $J_{\parallel}^{\text{sat}}$ rollover (e.g. figure 2 in Loarte *et al.* 1998) and electron temperature cliff (e.g. figure 5(a) in McLean *et al.* 2015) phenomena observed in detachment experiments as shown in figure 4. We admit that more sophisticated models such as 2D UEDGE and SOLPS-ITER considering additional details in principle yield more trustworthy results; but the fundamental physics picture of plasma detachment, that plasma radiates enough power to reach a low-enough temperature (a few electronvolts) and to recombine into neutral gas, will not change. Rather than developing a more comprehensive data-driven model with training data from state-of-the-art numerical simulations, here we focus on exploring the data-driven approach for detachment prediction with a somewhat simplified physics-based model (1D UEDGE) that could cover lots of plasma discharge scenarios relatively cheap computationally.

3.2. Data generation

Although there are more than a dozen independent geometry and plasma parameters that can be adjusted in the 1D UEDGE model, here we only vary the three most relevant parameters, namely, upstream density $n_{e,u}$, injection power P_{inj} and impurity fraction f_z . These three parameters ($n_{e,u}, P_{\text{inj}}, f_z$) are hereafter referred to as ‘model inputs’.

$P = 1.29\text{MW}$, $f_{rc} = 2.04\%$

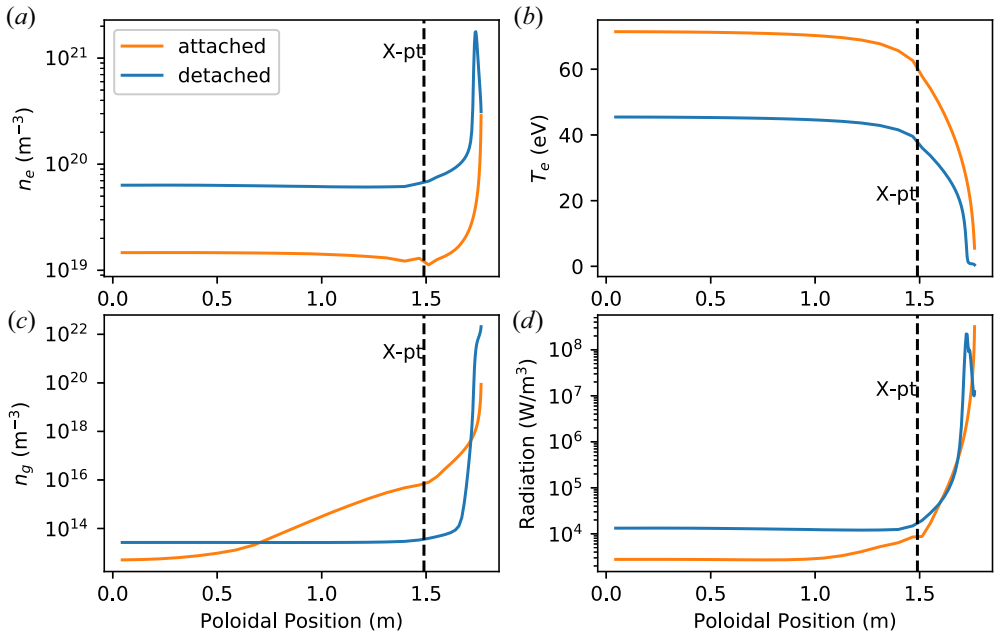


FIGURE 3. Examples of (a) plasma density, (b) electron temperature, (c) neutral density and (d) radiation profile of attached (blue) and detached (orange) divertor plasma from 1D UEDGE simulations.

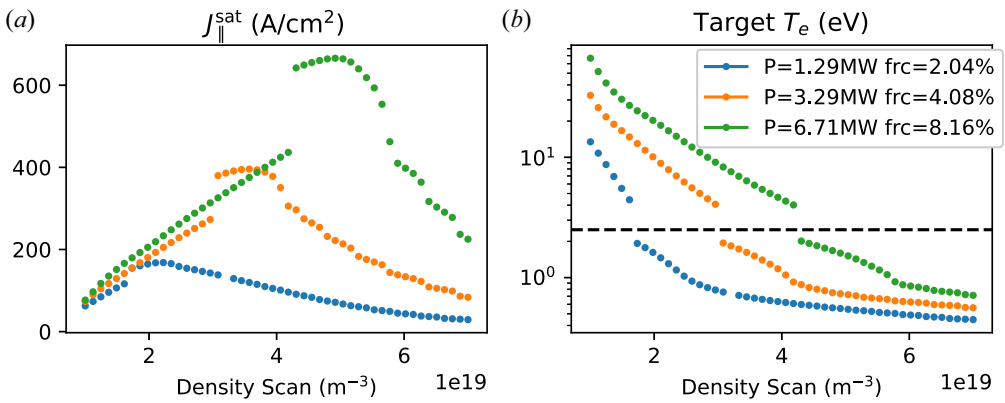


FIGURE 4. Reproductions of (a) $J_{\parallel}^{\text{sat}}$ rollover and (b) T_e cliff features in 1D UEDGE density scan simulations.

In this study, we use DIII-D tokamak geometry with a fixed divertor leg length $L_{\text{leg}} = 0.2696$ m, defined as the poloidal distance between the X-point to the outer divertor target as shown in figure 2(b). We uniformly sample $n_{e,u} \in [1, 7] \times 10^{19} \text{ m}^{-3}$, $P_{\text{inj}} \in [1, 10]$ MW and $f_Z \in 0\text{--}10\%$. Because the first wall material of DIII-D tokamak is graphite, here the impurity species is set to be carbon. With 60 sample points for $n_{e,u}$ and f_Z , and 40 sample points for P_{inj} , 144 000 cases in total are launched. However, not all of these cases can reach a physically feasible steady-state solution within a limited simulation time. Hence,

Data set	n_{total}	$n_{\text{detach}}(\%)$	$n_{\text{attach}}(\%)$
Training	111 598	56 047 (50.22)	55 551 (49.78)
validation	57 655	28 468 (49.38)	29 187 (50.62)

TABLE 1. Numbers of detached and attached cases and their percentages in the training and validation data sets.

only 111 598 converged simulation cases with good power balance relation are accepted as the training data set. Similarly, an independent validation data set with a coarser sampling rate ($50 \times 50 \times 30$ and 57 655 accepted solutions) is generated. Unless stated otherwise, the neural networks presented in this paper are trained and validated based on the training data set, only the combined forward prediction model is validated with the validation data set.

To avoid manual intervention of hundreds of thousands simulations individually, all the UEDGE runs are launched and managed by Merlin (Peterson *et al.* 2022), a ML workflow framework designed for large-scale high-performance computing (HPC) environments. It takes 4 days with 200 CPUs to obtain about 170 000 converged cases. The data-generation time could be further shortened if more CPUs were utilised.

For the diagnostic description of plasma state in this 1D set-up, five synthetic measurements are taken across the simulation domain, including the upstream electron temperature $T_{e,u}$, ion saturation current density $J_{\parallel}^{\text{sat}}$, electron density $n_{e,t}$ and temperature $T_{e,t}$ at divertor target, as well as the radiation profile P_{rad} . As illustrated in figure 2(b), $T_{e,u}$ mimics the mid-plane Thompson scattering measurement, $J_{\parallel}^{\text{sat}}$ mimics the Langmuir probe measurement, $n_{e,t}$ and $T_{e,t}$ come from either divertor Thompson scattering or Langmuir probe and P_{rad} is deduced from either bolometer or C-III emission. Note here $J_{\parallel}^{\text{sat}}$, $T_{e,u}$, $n_{e,t}$ and $T_{e,t}$ are scalars, whereas P_{rad} is a 1D profile. These diagnostic measurements are collected for all 111 598 cases and then normalised for model training.

Prior model training, all cases are labelled as either ‘detached’ or ‘attached’ because there is no ‘partial detached’ state in the 1D system. Here the choice of detachment criterion is straightforward. As shown in figure 5, electron temperature at the divertor target $T_{e,t}$ has two distinct distributions around $T_{e,t} = 2.1$ eV. The gap at 2.1 eV is caused by the temperature cliff phenomenon shown in figure 4(b). Therefore, 56 047 cases with $T_{e,t} < 2.1$ eV are labelled ‘detached’, whereas the remaining 55 551 cases are labelled as ‘attached’ cases. As shown in table 1, nearly half of the cases in the validation data set are labelled ‘detached’, similar to the percentage of ‘detached’ cases in the training set. In DIII-D experiment, the temperature cliff occurs around 5 eV, this apparent discrepancy is mainly due to our simplified 1D set-up. The temperature cliff in our simulations is consistent with the sonic transition point where the ion parallel velocity reaches sonic moving away from the target. A thorough investigation of temperature cliff in one dimension is beyond the scope of this paper and will be presented in a separated paper by Zhao *et al.* (2022). We would like to point out that labelling cases is primarily to better quantify the models’ accuracy in § 5; it does not affect the training process at all. In addition, the lower $T_{e,t} \sim 0.22$ eV limit shown in figure 5 is likely related to the atomic physics. When $T_e < 1$ eV, the ionisation rate decreases whereas the recombination rate increases dramatically. In this temperature range, electron collisionality is very high such that three-body recombination becomes dominant. This process moves plasma potential energy to electron thermal energy, stopping T_e decreasing further.

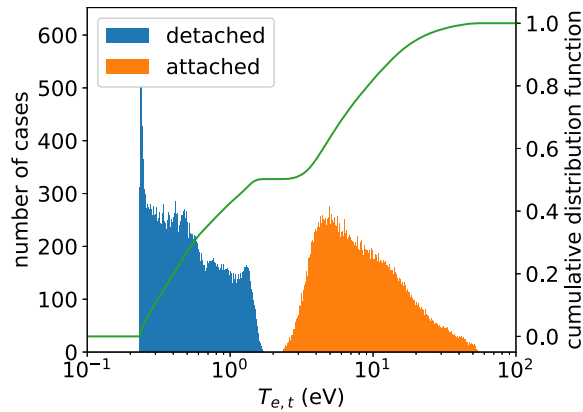


FIGURE 5. Histogram and cumulative distribution of electron temperature at divertor target $T_{e,t}$ from the training data set (111 598 cases in total).

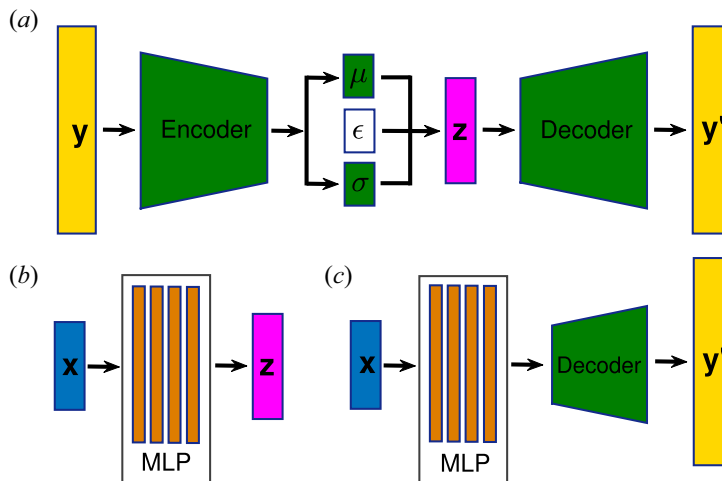


FIGURE 6. Sketches of (a) β -variational autoencoder, (b) forward multilayer perceptron (MLP) and (c) combined data-driven model.

4. Development of ML models

The goal of our ML-based approach is to predict the outputs of UEDGE simulations directly from a set of given model inputs, without actually performing UEDGE simulations and with significant speed up. As has been shown in the literature (Anirudh *et al.* 2020), this task becomes substantially easier and more tractable if the prediction is made through an intermediate and reduced dimensional space also learned through ML (see figures 1 and 6). Therefore, in this work, we utilise two ML models. The first ML model, a *variational autoencoder* (VAE), is used to generate an invertible mapping from diagnostic measurements to a reduced-dimensional ‘*latent space*’, whereas the second ML model, a *surrogate model*, is used to predict UEDGE outputs in this latent space, given the inputs to UEDGE.

4.1. Latent space identification using AEs

The basic idea behind latent space identification (also referred to as dimensionality reduction or data compression) is straightforward. Given an input data point, the goal of an AE is to reduce it to fewer dimensions such that the reduced representation allows the original input to be reconstructed as faithfully as possible. An AE comprises two parts: (1) an encoder that learns to meaningfully encode the data into a set of values, a compact representation of the input (i.e. latent coordinates); and (2) a decoder that attempts to reconstruct the original inputs from this latent space encoding. By forcing the reconstruction of data, an AE attempts to learn a suitable latent space that captures all necessary degrees of freedom while discarding trivial variations, noise and redundant correlations in the data.

In this work, we utilise two types of AE to create a suitable LSR in two steps. First, we develop a (vanilla) AE to identify an appropriate level of reduction, i.e. the dimensionality of the latent space and a suitable neural network architecture. Next, we turn this AE into a special type of AE, called the β -VAE, to create the final LSR.

4.1.1. AE design to identify the dimensionality of a suitable latent space

Given a set of diagnostic measurements (\mathbf{y}), we develop an AE to identify the compressed LSR (\mathbf{z}) of the plasma state. We use \mathbf{y}' to denote the reconstructed input from the AE. In other words, the encoder maps $\mathbf{y} \rightarrow \mathbf{z}$ and the decoder maps $\mathbf{z} \rightarrow \mathbf{y}'$. In our case, \mathbf{y} (and, hence, also \mathbf{y}') has 34 values: it comprises 4 scalar values ($J_{\parallel}^{\text{sat}}$, $T_{e,u}$, N_t and $T_{e,t}$) and a 30-element 1D array (P_{rad}). However, the different values (scalars versus the array) exhibit different ranges of values. Therefore, we utilise a scaling factor α_i for each of the 34 input dimensions to prevent overfitting on the 30 values for the array. In particular, we choose $\alpha_i = 0.1$ for values corresponding to the array elements and $\alpha_i = 1.0$ for the scalars, effectively asking the model to consider each array element $10\times$ less important than the scalars. Formally, our AE is trained to minimise the loss function

$$\mathcal{L}(\mathbf{y}, \mathbf{y}') = \frac{1}{N} \sum_i \alpha_i |y_i - y'_i|^n, \quad (4.1)$$

where $N = 34$ is the total number of elements in the input diagnostic measurement data \mathbf{y} . For $n = 1$, the loss function is the mean absolute error (MAE) or $L1$ norm, and for $n = 2$, it becomes the mean square error (MSE) or $L2$ norm.

As our input data are simply a collection of values, our AE design uses a series of fully connected neural network layers. Each such layer provides dense connections between input and output values through a linear transformation (a matrix multiplication with an additive bias) followed by a nonlinear activation. Given a chosen input/output dimensionality and an activation function, the training of the model then learns the appropriate matrix and bias to perform this transformation. We experimented with different depths of the neural network (i.e. the number of such fully connected layers) and the corresponding dimensionalities. All models were trained using the Tensorflow (Abadi *et al.* 2015) and Keras (Chollet *et al.* 2015) frameworks with the Adam optimiser (Kingma & Ba 2014). A dataset of 111 598 samples was used for developing the AE, with a randomly selected 80 % subset of these data for training, and the remaining 20 % for validating the model (i.e. assess its accuracy).

The final AE model, which provided the best reconstruction accuracy (see table 2), contains three fully connected layers that progressively reduce the data dimensionality as $34 \rightarrow 18 \rightarrow 10 \rightarrow 6$. Each layer is followed by the sigmoid activation function. The output of the last layer is the desired LSR; in other words, a 6-dimensional (6D) latent

D_z	8	7	6	5	4
$\mathcal{L}(\mathbf{y}, \mathbf{y}')^{n=1}$	3.5×10^{-4}	4.2×10^{-4}	4.7×10^{-4}	4.8×10^{-4}	5.2×10^{-4}
$\mathcal{L}(\mathbf{y}, \mathbf{y}')^{n=2}$	2.5×10^{-6}	3.1×10^{-6}	3.3×10^{-6}	4.5×10^{-6}	5.2×10^{-6}

TABLE 2. To choose a suitable dimensionality, D_z , of the latent space, we study the reconstruction errors posed by different dimensionalities. Given the significant jump in the error when going from six to five, D_z was found to be the appropriate choice.

space is adequate to compress a 34-element diagnostic data in our case so as to recover the input with sufficient accuracy.

4.1.2. VAE design to identify a suitable latent space

The AE set-up described here is capable of generating compressed encoding of the input data. However, such AEs provide no control of the distribution of the data in the learned LSR. Typically, this is not a big issue if the AE is used to encode/decode the data, i.e. use the AE (both encoder and decoder) as one model. However, it is not suitable for our application as our goal is to utilise only the decoder and in conjunction with a forward, surrogate model. Specifically, in our case it is important to provide smoothness guarantees on the LSR, otherwise even small prediction errors from the forward model could be amplified significantly by the decoder. To regularise the distribution in the latent space and to improve decoder's generative capability, we use a specific type of AE, called β -VAE (Higgins *et al.* 2016). The key difference between the (vanilla) AE and the VAE is that the VAE forces the distribution of \mathbf{z} to a given reference distribution, typically a multivariate normal distribution, offering some important mathematical guarantees, such as smoothness. To ensure that \mathbf{z} mimics the reference distribution, a regularisation term is added to the loss function to measure how different the two distributions are. Specifically, the loss function from (4.1) is modified to

$$\mathcal{L}(\mathbf{y}, \mathbf{y}') = \frac{1}{N} \sum_i \alpha_i |y_i - y'_i|^n + \beta D_{\text{KL}}(f(\mathbf{z}) || f_0), \quad (4.2)$$

where $D_{\text{KL}}(f(\mathbf{z}) || f_0)$ is the Kullback–Leibler divergence, which measures how the distribution $f(\mathbf{z})$ differs from the reference distribution f_0 , and a hyperparameter β is used to balance the decoder reconstruction accuracy (the first term in the loss function) and the orthogonality of latent space coordinates (the second term). As is common for VAEs, our reference distribution f_0 is a multivariate normal distribution, $\mathcal{N}(0, 1)$. In many applications, β is set to be larger than unity for better separation of independent latent space coordinate; in our case, forcing the LSR distribution function to match the normal distribution is not necessary. Instead, a small $\beta = 10^{-9}$ is found to be suitable. The VAE used in this study has the same network architecture and training parameters as the AE described previously.

4.2. Predicting simulation outputs using a surrogate model

Given the LSR, \mathbf{z} , generated by the VAE described previously, we next develop another ML model that can predict \mathbf{z} based on the physics model inputs, \mathbf{x} (i.e. global discharge parameters). Due to a small numbers of model inputs (i.e. three varying parameters in 1D UEDGE model) and outputs (i.e. 6D latent space), here a simple yet robust MLP is chosen to be our forward model. This MLP has four fully connected layers with 24

neurons on each layer, uses the rectified linear (ReLU) activation function, and MSE as the loss function. An Adam optimiser with a learning rate 0.001 is used for training the model.

5. Data-driven model performance

In this section, we discuss the results on the performance and validation of the ML models developed using the methods discussed in § 4.

5.1. Latent space identification using AEs

Through the AE described previously, we identify the dimensionality of a suitable latent space, D_z , to be six. Table 2 indicates that with both $L1$ or $L2$ norm, the validation loss (in normalised units) increases as D_z decreases. The $L2$ norm appears to have a critical turning point at $D_z = 6$; the validation loss increases substantially when D_z is set to be less than six, whereas for $D_z > 6$, the improvement is less profound.

Next, we study the performance of the VAE using figure 7 with UEDGE data f_{UEDGE} on the horizontal axis and the residual between VAE-reproduced data and UEDGE data $\varepsilon = f_{\beta\text{-VAE}} - f_{\text{UEDGE}}$ on the vertical axis. If a perfect reconstruction can be achieved, all data points would lie exactly on the horizontal $\varepsilon = 0$ lines. In our case, two of the four scalar diagnostic measurements, the electron density at divertor $n_{e,t}$ and the electron upstream temperature $T_{e,u}$ can be reproduced with excellent accuracy, whereas the other two scalar measurements, the ion saturation current, $J_{\parallel}^{\text{sat}}$, and the electron temperature at divertor, $T_{e,t}$, also show good quality reconstruction, with a slight performance degradation at high values of the respective diagnostics. Nevertheless, the R^2 scores (or, coefficient of determination that measures the goodness-of-fit between the two variables $f = \{f_i\}$, $g = \{g_i\}$ defined as $R^2 = 1 - \sum_i (f_i - g_i)^2 / \sum_i f_i^2$ with i being the index of the variable; hence, $R^2 = 1$ indicates perfect correlation whereas $R^2 = 0$ means no correlation) for these four scalars all exceed 0.997, evidencing a close to perfect replication of input data from trained β -VAE statistical-wise. To quantify the reconstruction quality of the radiation profile, P_{rad} , we choose two metrics: peak amplitude and location. The peak radiation amplitude prediction by the VAE is consistent the UEDGE data with a slightly larger variance ($R^2 \simeq 0.994 < 0.997$). It appears to be challenging to reproduce the peak radiation or detachment front location with $R_{\text{detached}}^2 \approx 0.7$ and $R_{\text{attached}}^2 \approx 0$. However, due to the discrete nature of the simulation mesh, radiation peak locations are discretised as well so that the R^2 score here could be misleading, especially for the attached cases. In fact, all 55 551 attached cases whose peaking radiation location at the divertor target are correctly reproduced by β -VAE but they are represented by what appears to be ‘a single’ orange dot in figure 7(f).

One plasma quantity that we are especially interested in for detachment prediction is the electron temperature at the divertor target, $T_{e,t}$, as this is arguably the most important and direct indicator of detachment. As shown in figure 8(a), VAE is able to reproduce $T_{e,t}$ near the detachment onset value (~ 2.1 eV for this study) very well. Figure 8(b) shows that the VAE is able to also accurately retain the ion exit velocity, $v_{\parallel i,e} = J_{\parallel}^{\text{sat}} / (en_{e,t})$, which is an inherent quantity that was not trained directly. A discrepancy between the VAE’s predictions and UEDGE results appears for $J_{\parallel}^{\text{sat}} / (en_{e,t}) > 55 \text{ km s}^{-1}$, which corresponds to attached plasma cases with $T_{e,t} \gg 2.1$ eV. However, this type of discrepancy exists only for less than 2% of the total cases: the small number of $T_{e,t} \gg 2.1$ eV samples may cause the relatively large prediction error for these cases. This result indicates that our trained VAE captures the correct physics constraints between different plasma variables (i.e. AE inputs and outputs) for at least the majority of the cases in our dataset.

Finally, we show that β -VAE encoded LSRs are indeed apart in latent space based on plasma states (e.g. detached or attached). Figure 9 illustrates the distinctive clusters

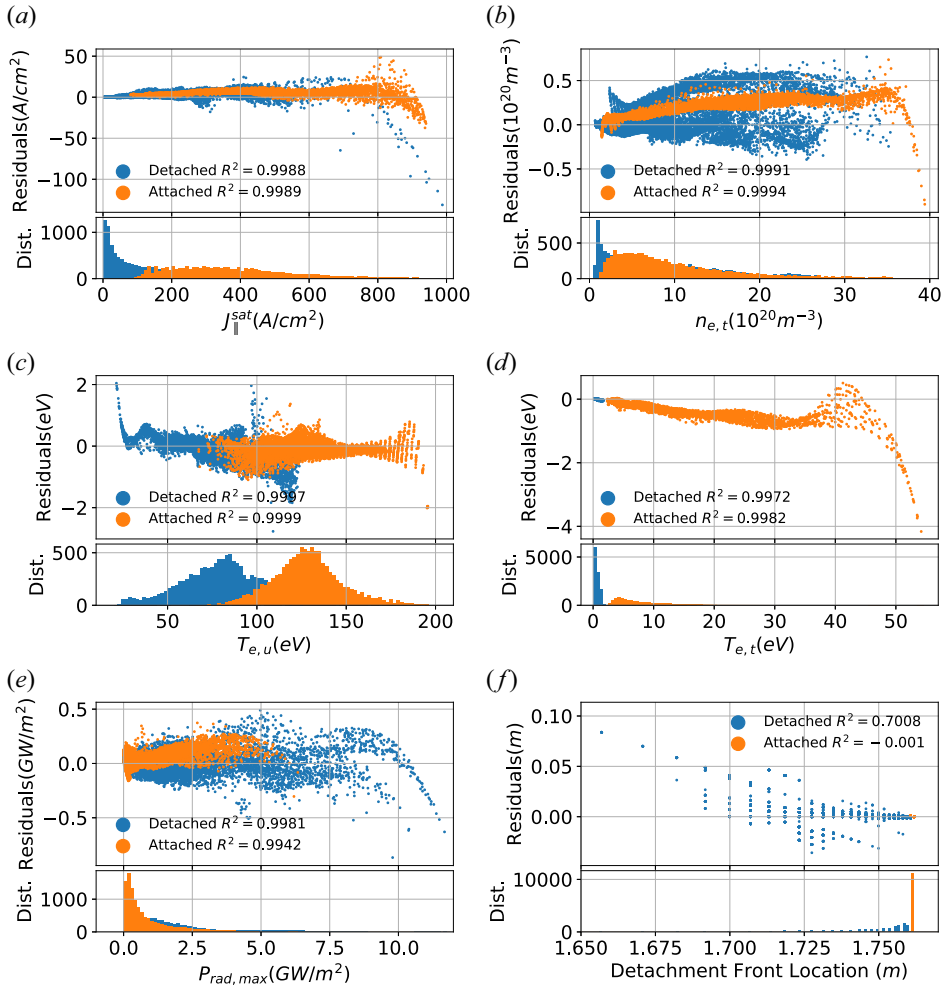


FIGURE 7. Performance of β -VAE in terms of absolute residual error for (a) $J_{\parallel}^{\text{sat}}$, (b) $n_{e,t}$, (c) $T_{e,u}$, (d) $T_{e,t}$, (e) $P_{\text{rad,max}}$ and (f) peak radiation or detachment front location. Here, the x -axis is UEDGE (true) value and the residual error is defined as $\varepsilon = f_{\beta\text{-VAE}} - f_{\text{UEDGE}}$ for the quantity f . All predictions show excellent correlation with the true values, as quantified with almost-perfect R^2 scores. The distribution of data points are shown in adjoining panels to overcome the issue of overplotting in the scatter plots; these indicate that for most cases, relatively low and unbiased error is obtained, whereas higher errors are associated with relatively smaller number of samples.

of LSRs for detached and attached cases in latent space using the t -distributed stochastic neighbour embedding (t -SNE) method (Van der Maaten & Hinton 2008). Similarly, figure 10 has the distributions of all six latent coordinates, again separated by attachment/detachment label. The result shows that despite overlaps between detached and attached plasmas in some latent variables (e.g. latent variables 1, 4 and 5), the two scenarios are broadly well separated in the other latent variables (e.g. latent variables 2 and 6), which is a nice feature to have for the upcoming forward detachment prediction.

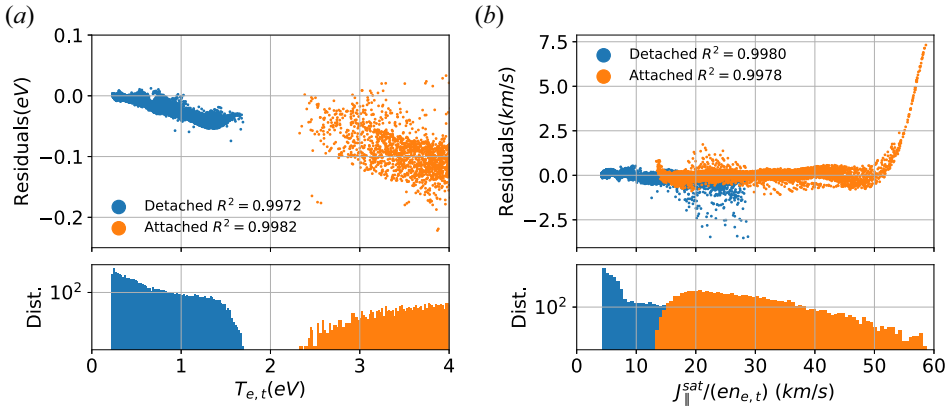


FIGURE 8. Residual plots for (a) $T_{e,t}$ near detachment transition and (b) $J_{\parallel}^{\text{sat}}/n_{e,t}$. Detached cases are in blue whereas attached cases are in orange. Note here that the y-axis of the distribution plots is in logarithmic scale.

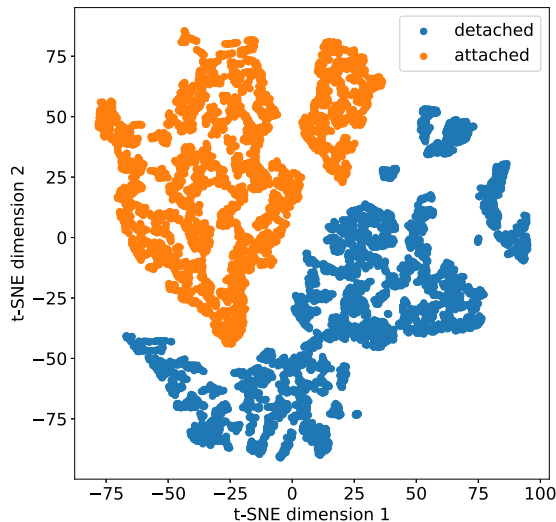


FIGURE 9. t -SNE visualisation of a subset of 10 000 randomly picked LSR samples in the training data set in which detached (blue) and attached (orange) cases are clearly separated.

5.2. Predicting simulation outputs using a surrogate model

Here, we study the performance of our surrogate model, an MLP, after 10 000 epochs. In particular, we compare the LSR generated by the VAE, z , and that predicted by the surrogate model, z_f . Figure 10 depicts plots the six coordinates within the two quantities (z vs. z_f) as individual scatter plots and shows high degree of correlation. As shown in the figure, the R^2 values for all six coordinates exceeds 0.99, indicating exceptional performance and that our surrogate model is capable of predicting the LSR with almost full accuracy.

5.3. Forward prediction

With both β -VAE and forward model trained, we are now able to predict diagnostic measurement of a plasma state from the model inputs by combining the forward model

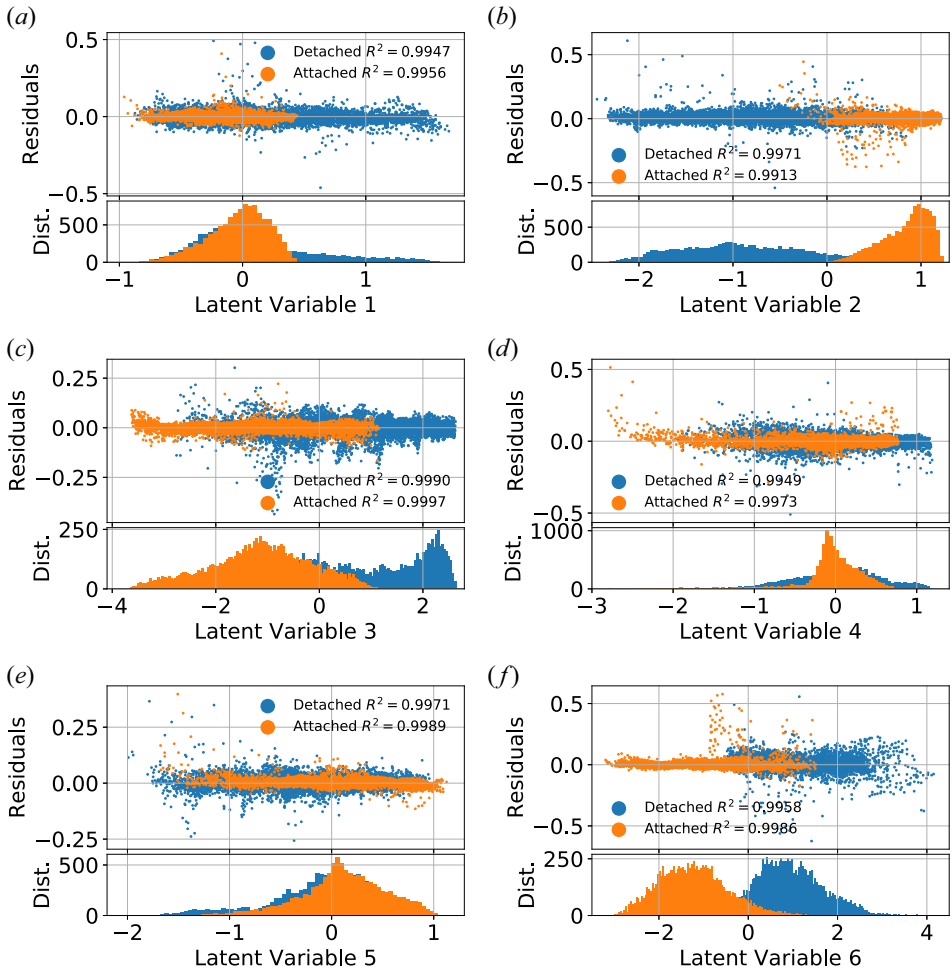


FIGURE 10. Performance of the MLP model in terms of absolute residual error for all six latent variables. Here the x -axis is β -VAE encoded value, whereas the residual is defined as $\varepsilon = f_{\text{MLP}} - f_{\beta\text{-VAE}}$ for quantity f . As with earlier results, an almost-perfect R^2 score indicates exceptional prediction quality.

and decoder (figure 6c). To ensure that this model is evaluated properly, a separated validation data set consists 57 655 cases generated independently following the similar data-generation process described in § 3.2.

The performance of forward detachment prediction model is illustrated in figure 11 where all 57 655 cases are evaluated with the ML model then validated with the UEDGE simulation results. As expected, the overall accuracy degrades marginally compared with the performance of β -VAE (e.g. figure 7) in this sequential MLP and decoder architecture.

Figure 12 and table 3 elucidate the accuracy of this combined forward detachment prediction model, quantified with error statistic analysis. We examine model accuracy for both detached and attached cases, and find that our model does perform slightly different for different plasma states. For attached plasma, it gives better results when predicting certain plasma quantities, such as ion saturation current and divertor target electron density, whereas it is less satisfactory for divertor target electron temperature

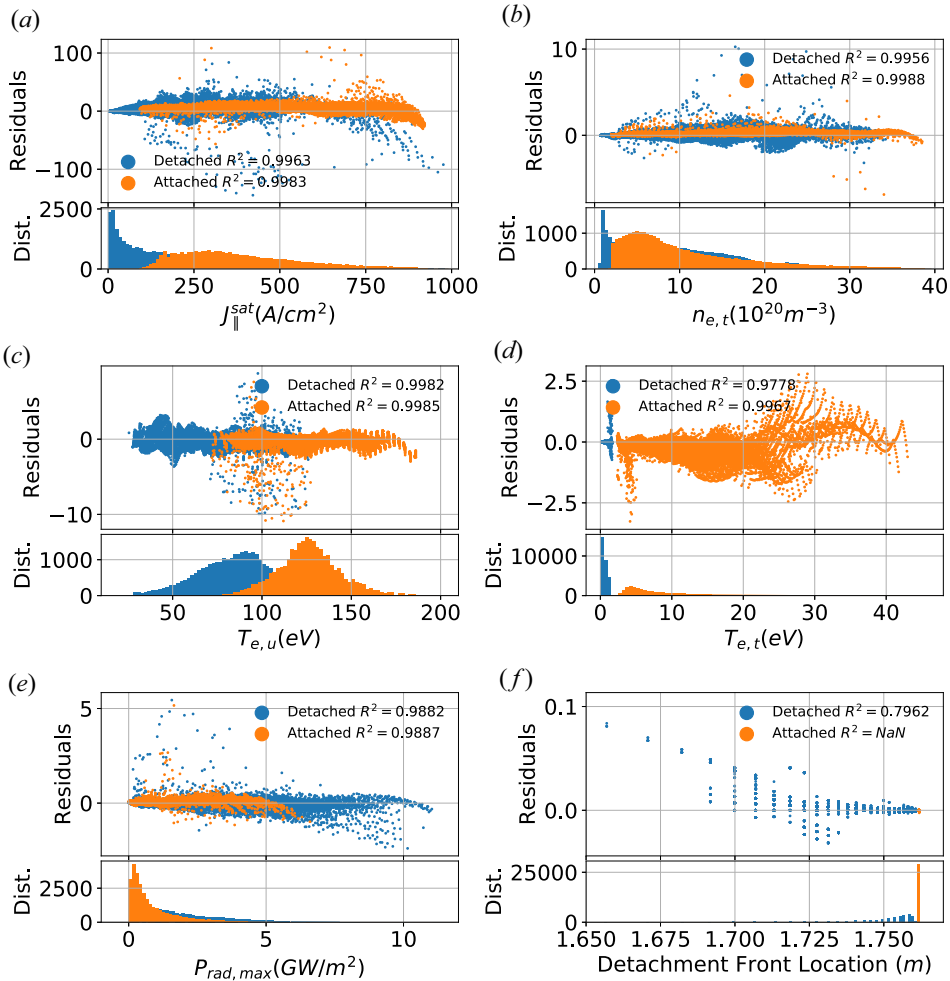


FIGURE 11. Performance of combined forward prediction model in terms of residuals for (a) $J_{\parallel}^{\text{sat}}$, (b) $n_{e,t}$, (c) $T_{e,u}$, (d) $T_{e,t}$, (e) $P_{\text{rad,max}}$ and (f) peak radiation or detachment front location. Here the x -axis is UEDGE (true) value, whereas the residual is defined as $f_{\text{residual}} = f_{\text{ML}} - f_{\text{UEDGE}}$ for quantity f . The distribution of f and the R^2 scores between f_{ML} and f_{UEDGE} for detached (blue) and attached (orange) cases are also provided.

prediction. However, no significant quantitative difference is found except the detachment front location prediction. Out of four scalar measurements, upstream temperature $T_{e,u}$ is the most accurately predicted diagnostic quantity with mean and standard derivation of relative error $\mu = -0.37\%$ and $\sigma = 0.83$ for the entire validation data set. The other three scalar measurements are also well predicted with $|\mu| < 3\%$ and $\sigma < 4\%$ for all the relative errors. The peak radiation strength prediction appears to have a fairly large uncertainty ($\sim 10\%$). This is possibly due to the shape peak structure (i.e. lack of resolution) for detached plasmas and boundary effect (i.e. maximum value resides at the last point) for attached plasmas as shown in figure 13(b). Fortunately, the more useful information, the peak radiation location (equivalent to the detachment front location L_{DF} to some extent) prediction is remarkably accurate. Here L_{DF} is correctly predicted to be at the divertor target plate for nearly all attached plasmas, and has a small uncertainty

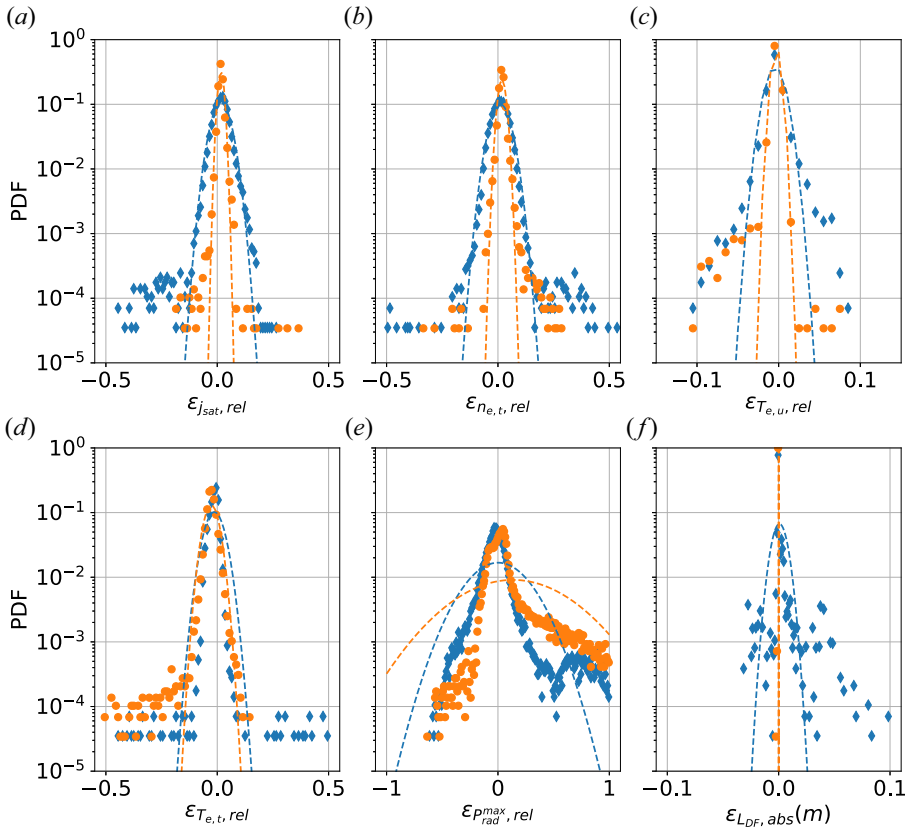


FIGURE 12. Probability distribution functions (PDFs) of relative error of (a) $J_{\parallel}^{\text{sat}}$, (b) $n_{e,t}$, (c) $T_{e,u}$, (d) $T_{e,t}$, (e) $P_{\text{rad}}^{\text{max}}$ and (f) absolute error of peak radiation or detachment front location. Detached cases are in blue whereas attached cases are in orange. Dashed lines are shifted Gaussian distribution functions with μ and σ from table 3 correspondingly.

($\sigma = 0.60$ cm) for detached plasmas. Note that the error distributions of four scalar measurements appear to follow Gaussian distribution, whereas for the peak radiation amplitude and location, the error distribution are no longer Gaussian: the standard derivation σ is skewed by rare (i.e. probability $\sim 0.1\%$) extreme cases.

We are particularly interested in $T_{e,t}$ prediction as this is the primary indicator of detachment. If we use the same detachment criterion (e.g. $T_{e,t} = 2.1$ eV), there are 35 out of 28 468 detached cases (i.e. blue triangles at the top left quadrant of figure 13a) misclassified as ‘attached’. Similarly, 25 out of 29 187 attached cases (orange diamonds at the bottom right quadrant of figure 13a) are mislabelled ‘detached’. Even counting in the other marginal cases, the misclassification rate is lower than 0.2%. These misclassified cases appears to congregate near the detachment onset point (i.e. all these cases have $T_{e,t} \in (1, 4)$ eV). However, no common feature can be identified in terms of the control parameters x .

In addition to accuracy, speed is another important metric for a real-world application. Table 4 summarises our predictive model’s performance in terms of the wall-clock time required to predict diagnostic measurements based on controlled inputs. Even without any optimisation, this model takes about 36 ms to carry out a prediction for one case: about 10 000 times faster than 1D UEDGE simulations which normally require a few minutes to

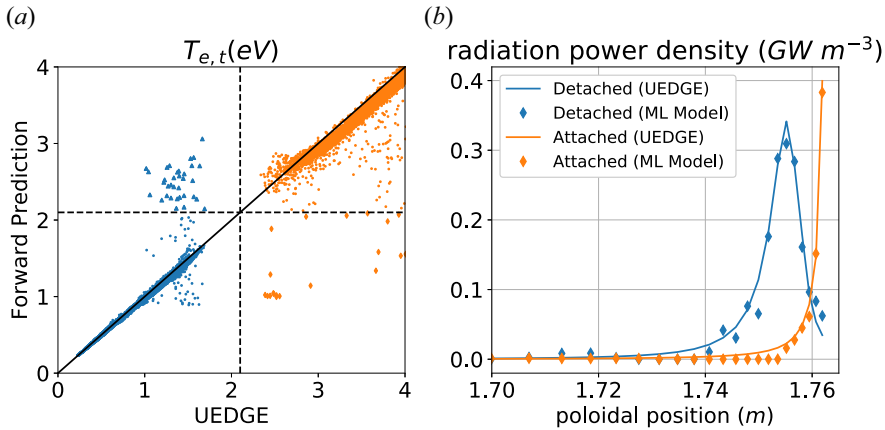


FIGURE 13. (a) Electron temperature at divertor target $T_{e,t}$ prediction versus UEDGE simulation result and (b) examples of radiation profile or detachment front prediction. True (UEDGE) detached cases are in blue whereas attached cases are in orange.

	Detached		Attached		Combined	
	μ	σ	μ	σ	μ	σ
$\epsilon_{J_{\parallel}^{\text{sat}},\text{rel}}(\%)$	1.66	3.76	1.67	1.26	1.67	2.79
$\epsilon_{n_{e,t},\text{rel}}(\%)$	1.06	3.97	1.81	1.66	1.44	3.05
$\epsilon_{T_{e,t},\text{rel}}(\%)$	-0.43	1.05	-0.30	0.52	-0.37	0.83
$\epsilon_{T_{e,t},\text{rel}}(\%)$	-0.13	3.90	-2.70	3.06	-2.01	3.57
$\epsilon_{P_{\text{rad}}^{\text{max}},\text{rel}}(\%)$	0.19	23.76	13.39	43.85	6.88	36.00
$\epsilon_{L_{\text{DF}},\text{abs}}(\text{cm})$	0.07	0.60	0	0	0.04	0.42

TABLE 3. Mean (μ) and standard deviation (σ) of the relative or absolute error of forward prediction model.

Number of cases	1	10	100	1000	10 000	100 000
Speed (ms)	36	37	42	51	178	2259

TABLE 4. Wall-clock time versus number of cases.

find a converged solution. It is already within the minimum requirement for detachment control (~ 100 ms). The model efficiency also increases when predicting multiple cases (for $n < 100\,000$).

6. Comparison with 2PMs

Validation of the forward model for detachment prediction shows that our model is able to accurately predict UEDGE result with a dramatic speed-up, but perhaps a more meaningful benchmark exercise would be comparing our newly developed model, as well as the UEDGE model, with the widely used detachment prediction models nowadays,

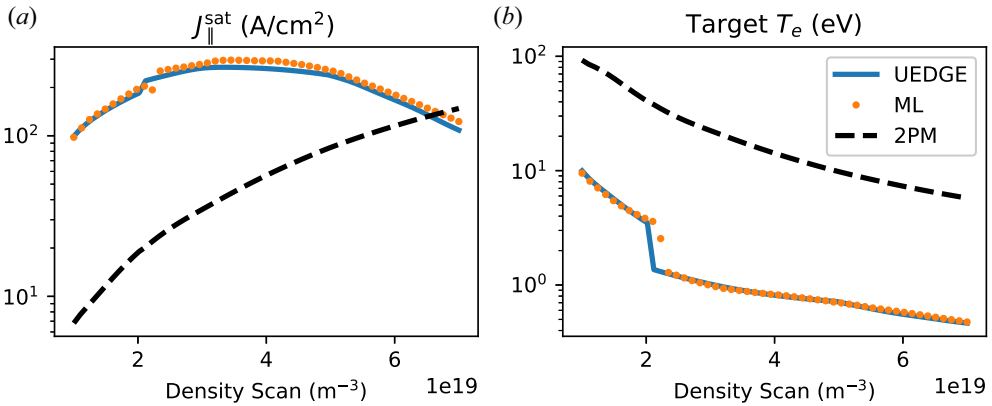


FIGURE 14. Upstream density scan of (a) $J_{\parallel}^{\text{sat}}$ and (b) $T_{e,u}$ for UEDGE (blue), ML data-driven model (orange) and basic 2PM (black dashed) with $P_{\text{inj}} = 1.71$ MW and $f_z = 0$.

namely the analytical basic 2PM (Stangeby *et al.* 2000) and the most sophisticated semi-analytical 2PMF (Stangeby 2018).

6.1. Basic 2PM

The basic 2PM is derived to evaluate $T_{e,u}$, $T_{e,t}$ and $n_{e,t}$ for given $n_{e,u}$, parallel heat flux q_{\parallel} and flux tube length L based on particle, pressure and power balance. Because of its simplicity, basic 2PM has been implemented in the tokamak detachment control *a priori* to estimate the degree of detachment (Eldon *et al.* 2022). Because 2PM does not account for momentum and power loss between the upstream and downstream points, we therefore benchmark UEDGE, the forward detachment model and the 2PM on a case with zero impurity to remove the impurity radiation. Figure 14 displays upstream density scan of $J_{\parallel}^{\text{sat}}$ and $T_{e,t}$ for the three models. Clearly, the data-driven model prediction and the UEDGE simulation results are in good agreement whereas the 2PM gives a quite different result. This one-order-of-magnitude discrepancy between our data-driven model/UEDGE and 2PM is due to the more comprehensive physics in our model/UEDGE. For instance, in this comparison the upstream ion temperature is much higher than the electron temperature in the upstream region in the UEDGE simulations due to slow ion parallel thermal conduction whereas the electron and ion temperatures are assumed to be the same in the basic 2PM. In addition, in the basic 2PM, the parallel heat flux is assumed to be conductive using the Spitzer–Härm formula, whereas in 1D UEDGE simulations, the parallel heat flux is assumed to be local flux-limited thermal transport. UEDGE is only able to recover 2PM result by further simplifying simulation (e.g. using slab configuration and turn off non-ideal terms) and manually adjusting boundary conditions to match 2PM assumptions, such as $T_e = T_i$ upstream.

6.2. 2PMF

To address the volumetric momentum and power losses, and the magnetic geometry effects, the basic 2PM has been extended to semi-analytical 2PMF with pre-fitted power and momentum loss coefficients f_{cooling} , $f_{\text{mom-loss}}$ (Stangeby 2018). Benchmark between our data-driven model, UEDGE simulation and 2PMF are also performed. Two fitting curves (from Stangeby 2018) are proposed for both power and momentum loss coefficients in 2PMF:

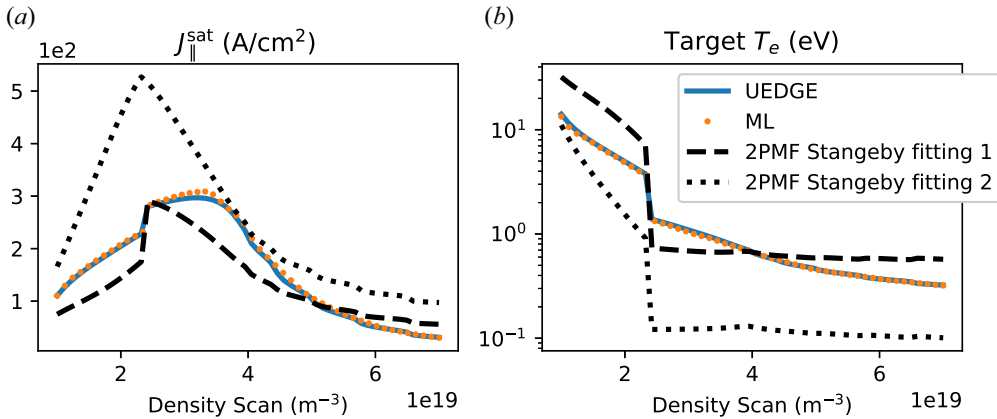


FIGURE 15. Upstream density scan of (a) $J_{\parallel}^{\text{sat}}$ and (b) $T_{e,u}$ for UEDGE (blue), ML data-driven model (orange) and 2PMF (black lines) with $P_{\text{inj}} = 2.66$ MW and $f_Z = 0.02$.

Stangeby fitting formula 1,

$$1 - f_{\text{cooling}} = [1 - \exp(-T_{e,t}/2.4)]^{1.9} \quad (6.1)$$

$$1 - f_{\text{mom-loss}} = [1 - \exp(-T_{e,t}/0.8)]^{2.1}; \quad (6.2)$$

Stangeby fitting formula 2,

$$1 - f_{\text{cooling}} = 0.9[1 - \exp(-T_{e,t}/6)]^{1.7} \quad (6.3)$$

$$1 - f_{\text{mom-loss}} = 1.3[1 - \exp(-T_{e,t}/1.8)]^{1.6}; \quad (6.4)$$

where UEDGE observed $T_{e,t}$ is used to estimate f_{cooling} and $f_{\text{mom-loss}}$. Figure 15 displays upstream density scan of $J_{\parallel}^{\text{sat}}$ and $T_{e,t}$ for the three models. The results from 2PMF compare with UEDGE simulation results better than those from the basic 2PM. All three models show two common features: (1) the ion saturation current density rollover and (2) temperature cliff at the onset of detachment. Once again, our data-driven model prediction matches UEDGE simulation results very well. Although there is no qualitative disagreement between our data-driven model/UEDGE and 2PMF, quantitatively the predictions between data-driven model/UEDGE and 2PMF can be off by an order of magnitude. This is because 2PMF prediction depends heavily on the choice of fitting curves for the power and momentum loss coefficients which may simplify the nonlinear dynamics setting the divertor plasma conditions. As illustrated in figure 15, fitting formula 1 does a better overall job than fitting formula 2 for this test case as $T_{e,t}$ predicted by fitting formula 2 is nearly an order of magnitude lower than UEDGE result when divertor plasma is detached ($n_{e,u} > 2.4 \times 10^{19} \text{ m}^{-3}$).

7. Model applications

As mentioned in the introduction, the main motivation of developing such a surrogate model is to enable reliable and fast detachment prediction for integrated machine design, scenario development and real-time plasma control. Heat exhaust is not a severe issue for current tokamaks due to the overall limited power output. Therefore, the SOL and divertor plasma dynamics is less of a concern and most of the effort so far has been focused on exploring operation scenarios with improved fusion performance,

equilibria and plasma stability control inside the separatrix, e.g. core–edge integration. However, for future high-power fusion devices such as reactors, their operation space must also fulfill constraints posed by divertor’s material and engineering limits. Likewise, fusion burn control needs to incorporate divertor heat and particle exhaust solutions such as detachment. In other words, designing and operating future devices require core–edge–SOL/divertor integration. SOL/divertor modelling suffers from the accuracy–speed trade-off similar to many other research. Current SOL/divertor transport codes such as UEDGE and SOLPS, are too slow for these applications as they are designed for physics investigations, whereas the basic 2PM and 2PMF are fast but perhaps over-simplify the problem. Our data-driven model overcomes this accuracy–speed gap with some room to prioritise one factor over another depending on the application. For instance, speed is the top criterion for plasma control (either in a simulator or in the actual plasma control system). Real-time or even faster than real-time prediction is required in order to activate actuators in time. On the other hand, accuracy is likely weighted more than speed for device design and scenario development.

The accuracy of the proposed data-driven approach relies on the quality and quantity of the training data set. We remark that tokamak edge plasma contains very rich physics with many factors have influence on the detachment onset or threshold, such as 2D/three-dimensional (3D) effects, divertor plate geometry, multi-species and/or multi-charge-state impurity and wall condition. Owing to the 1D flux-tube mesh simplification, 1D UEDGE simulations under-estimate the detachment onset temperature $T_{e,t}$ and over-estimate the peak radiation amplitude, our current model unsurprisingly picks up these unfavourable predictions. One would expect that the performance of the model improves in terms of matching real experimental measurement when the training data are extended to incorporate richer physics in a more realistic experiment setting, i.e. once trained upon higher-quality data sets either from higher-fidelity numerical models such as 2D UEDGE/SOLPS-ITER or experiments. Even though the underlying methodology will be the same, the architecture of neural networks likely needs to be modified. Notably, synthetic or real diagnostic measurements will have a complicated format with 2D simulations or experiments. In a realistic tokamak, both Langmuir probe and Thompson scattering are multi-channel; therefore, $J_{\parallel}^{\text{sat}}$, $T_{e,u}$, $n_{e,t}$ and $T_{e,t}$ become sparse 1D arrays in space (point data), spectroscopic and bolometer diagnostics provide radiation power/strength at certain wavelength or over the entire spectrum (e.g. 1D volume-averaged data), visible and infrared (IR) cameras give image (i.e. 2D data with projected volume and certain range of wavelength averaged quantity). Moreover, there are likely some discrepancies between model produced synthetic diagnostic data and real experimental measurements due to various reasons such as model simplification and inherent instrumentation noise. Handling these multi-modal diagnostics together in a consistent manner, as well as bridging the gap between simulation and experimental data will be addressed in our future study.

Even though the data-driven detachment prediction model presented here is based on 1D UEDGE simulations, it could be readily integrated for detachment control application. At first glance, current model is similar to the analytical basic 2PM which has recently been implemented in KSTAR tokamak detachment control system (Eldon *et al.* 2022) as both models are meant to predict divertor or downstream plasma state. Compared with the basic 2PM, our model features additional detachment front prediction. In addition, owing to fewer simplifications made in the UEDGE model such as B variation, local flux-limited thermal transport and impurity radiation effects, our model should give more reliable predictions than 2PM in certain circumstances, resulting improved control performance of detachment and overall plasma confinement.

8. Summary

In this paper, we explore a new physics model-based approach to predict divertor detachment by leveraging the ‘latent feature space’ concept in ML research. As a proof of concept study, a highly efficient 1D UEDGE model which contains the crucial physics ingredients of detachment is used to simulate the plasma and neutrals along the open magnetic field lines in the SOL and to generate our training and validation data sets. Over 160 000 simulations with 3 varying UEDGE model inputs \mathbf{x} (e.g. different upstream density, injection power and carbon fraction) are performed to cover the normal DIII-D tokamak operation parameter region; and 5 synthetic diagnostic measurements such as upstream temperature, electron density, temperature and saturation current at divertor target, as well as radiation profile are collected as the diagnostic set \mathbf{y} . The latent space as well as the LSR of plasma state \mathbf{z} are then identified by compressing \mathbf{y} through an AE. Sequentially, a forward surrogate model is trained to make predictions of \mathbf{z} from UEDGE model inputs \mathbf{x} ; then the trained decoder is used to reconstruct diagnostic measurement \mathbf{y} back in configuration space. We find that a 6D latent space is good enough to closely yield a match for the true system in configuration space (i.e. the synthetic diagnostic measurements), and the forward detachment prediction model ($\mathbf{x} \rightarrow \mathbf{z} \rightarrow \mathbf{y}$) also produces quite accurate predictions (relative error on the order of a few per cent statistically) with at least 10^4 speed-up compared with the UEDGE simulations.

Our pilot study demonstrates that the complicated divertor/SOL plasma state has a low-dimensional representation in latent space. Therefore, this new latent space description of plasma state can be used to not only construct a fast and robust surrogate model for steady-state detachment prediction as revealed in this paper, but also has the potential to be used for dynamical control once the critical plasma nonlinear dynamics (in latent space) was identified successfully.

Acknowledgements

The authors thank the anonymous reviewers for their comments and suggestions.

Editor William Dorland thanks the referees for their advice in evaluating this article.

Funding

This work is performed under the auspices of the U.S. Department of Energy (US DOE) by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 through the pilot project ‘A critical step for development of machine learning surrogate models for tokamak divertor-plasma detachment control’ (LLNL-JRNL-836030).

Declaration of interests

The authors report no conflict of interest.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ABADI, M., AGARWAL, A., BARHAM, P., BREVDO, E., CHEN, Z., CITRO, C., CORRADO, G.S., DAVIS, A., DEAN, J., DEVIN, M., *et al.* 2015 TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

- ANIRUDH, R., ARCHIBALD, R., ASIF, M.S., BECKER, M.M., BENKADDA, S., BREMER, P.-T., BUDÉ, R.H.S., CHANG, C.S., CHEN, L., CHURCHILL, R.M., *et al.* 2022 Review of data-driven plasma science. [arXiv:2205.15832](https://arxiv.org/abs/2205.15832).
- ANIRUDH, R., THIAGARAJAN, J.J., BREMER, P.-T. & SPEARS, B.K. 2020 Improved surrogates in inertial confinement fusion with manifold and cycle consistencies. *Proc. Natl Acad. Sci. USA* **117** (18), 9741–9746.
- BALLINGER, S.B., KUANG, A.Q., UMANSKY, M.V., BRUNNER, D., CANIK, J.M., GREENWALD, M., LORE, J.D., LABOMBARD, B., TERRY, J.L., WIGRAM, M., *et al.* 2021 Simulation of the SPARC plasma boundary with the UEDGE code. *Nucl. Fusion* **61** (8), 086014.
- VAN DEN BERG, J., ABRAMOVIC, I., LOPES CARDOZO, N.J. & MOSEEV, D. 2018 Fast analysis of collective Thomson scattering spectra on Wendelstein 7-X. *Rev. Sci. Instrum.* **89** (8), 083507.
- CHOLLET, F. & OTHERS 2015 Keras. <https://keras.io>.
- DEGRAVE, J., FELICI, F., BUCHLI, J., NEUNERT, M., TRACEY, B., CARPANESE, F., EWALDS, T., HAFNER, R., ABDOLMALEKI, A., DE LAS CASAS, D., *et al.* 2022 Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602** (7897), 414–419.
- DUDSON, B.D., ALLEN, J., BODY, T., CHAPMAN, B., LAU, C., TOWNLEY, L., MOULTON, D., HARRISON, J. & LIPSCHULTZ, B. 2019 The role of particle, energy and momentum losses in 1D simulations of divertor detachment. *Plasma Phys. Control. Fusion* **61** (6), 065008.
- ELDON, D., ANAND, H., BAK, J.G., BARR, J.L., HAHN, S.-H., JEONG, J.H., KIM, H.-S., LEE, H.H., LEONARD, A.W., SAMMULI, B.S., *et al.* 2022 Enhancement of detachment control with simplified real-time modelling on the KSTAR tokamak. *Plasma Phys. Control. Fusion* **64** (7), 075002.
- HIGGINS, I., MATTHEY, L., PAL, A., BURGESS, C., GLOROT, X., BOTVINICK, M., MOHAMED, S. & LERCHNER, A. 2016 beta-VAE: learning basic visual concepts with a constrained variational framework. In *ICLR 2017 Conference*.
- KATES-HARBECK, J., SVYATKOVSKIY, A. & TANG, W. 2019 Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568** (7753), 526–531.
- KINGMA, D.P. & BA, J. 2014 Adam: a method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- KUANG, A.Q., BALLINGER, S., BRUNNER, D., CANIK, J., CREELY, A.J., GRAY, T., GREENWALD, M., HUGHES, J.W., IRBY, J., LABOMBARD, B., *et al.* 2020 Divertor heat flux challenge and mitigation in SPARC. *J. Plasma Phys.* **86** (5).
- LOARTE, A., MONK, R.D., MARTIN-SOLIS, J.R., CAMPBELL, D.J., CHANKIN, A.V., CLEMENT, S., DAVIES, S.J., EHRENBERG, J., ERENTS, S.K., GUO, H.Y., *et al.* 1998 Plasma detachment in JET Mark I divertor experiments. *Nucl. Fusion* **38** (3), 331.
- MA, C., ZHU, B., XU, X.-Q. & WANG, W. 2020 Machine learning surrogate models for Landau fluid closure. *Phys. Plasmas* **27** (4), 042502.
- MATHEWS, A., FRANCISQUEZ, M., HUGHES, J.W., HATCH, D.R., ZHU, B. & ROGERS, B.N. 2021 Uncovering turbulent plasma dynamics via deep learning from partial observations. *Phys. Rev. E* **104** (2), 025205.
- MAULIK, R., GARLAND, N.A., BURBY, J.W., TANG, X.-Z. & BALAPRAKASH, P. 2020 Neural network representability of fully ionized plasma fluid model closures. *Phys. Plasmas* **27** (7), 072106.
- MAURIZIO, R., ELMORE, S., FEDORCZAK, N., GALLO, A., REIMERDES, H., LABIT, B., THEILER, C., TSUI, C.K., VIJVERS, W.A.J., TEAM, T.C.V., *et al.* 2017 Divertor power load studies for attached L-mode single-null plasmas in TCX. *Nucl. Fusion* **58** (1), 016052.
- MCLEAN, A.G., LEONARD, A.W., MAKOWSKI, M.A., GROTH, M., ALLEN, S.L., BOEDO, J.A., BRAY, B.D., BRIESEMEISTER, A.R., CARLSTROM, T.N., ELDON, D., *et al.* 2015 Electron pressure balance in the SOL through the transition to detachment. *J. Nucl. Mater.* **463**, 533–536.
- MILLER, M.A., CHURCHILL, R.M., DENER, A., CHANG, C.S., MUNSON, T. & HAGER, R. 2021 Encoder–decoder neural network for solving the nonlinear Fokker–Planck–Landau collision operator in XGC. *J. Plasma Phys.* **87** (2).
- MONTES, K.J., REA, C., GRANETZ, R.S., TINGUELY, R.A., EIDIETIS, N., MENEGHINI, O.M., CHEN, D.L., SHEN, B., XIAO, B.J., ERICKSON, K., *et al.* 2019 Machine learning for disruption warnings on Alcator C-Mod, DIII-D, and EAST. *Nucl. Fusion* **59** (9), 096015.

- PETERSON, J.L., BAY, B., KONING, J., ROBINSON, P., SEMLER, J., WHITE, J., ANIRUDH, R., ATHEY, K., BREMER, P.-T., DI NATALE, F., *et al.* 2022 Enabling machine learning-ready HPC ensembles with Merlin. *Future Gener. Comp. Sy.* **131**, 255–268.
- PORTER, G.D., ALLEN, S.L., BROWN, M., FENSTERMACHER, M.E., HILL, D.N., JONG, R.A., LEONARD, A.W., NILSON, D., RENSINK, M.E., ROGNLIEN, T.D., *et al.* 1996 Simulation of experimentally achieved DIII-D detached plasmas using the UEDGE code. *Phys. Plasmas* **3** (5), 1967–1975.
- REA, C., MONTES, K.J., PAU, A., GRANETZ, R.S. & SAUTER, O. 2020 Progress toward interpretable machine learning-based disruption predictors across tokamaks. *Fusion Sci. Technol.* **76** (8), 912–924.
- ROGNLIEN, T.D., BROWN, P.N., CAMPBELL, R.B., KAISER, T.B., KNOLL, D.A., MCHUGH, P.R., PORTER, G.D., RENSINK, M.E. & SMITH, G.R. 1994 2-D fluid transport simulations of gaseous/radiative divertors. *Contrib. Plasm. Phys.* **34** (2–3), 362–367.
- ROGNLIEN, T.D. & RENSINK, M.E. 2002 Edge-plasma models and characteristics for magnetic fusion energy devices. *Fusion Engng Des.* **60** (4), 497–514.
- SAMUEL, C.M., MCLEAN, A.G., JOHNSON, C.A., GLASS, F. & JAERVINEN, A.E. 2021 Measuring the electron temperature and identifying plasma detachment using machine learning and spectroscopy. *Rev. Sci. Instrum.* **92** (4), 043520.
- STANGEBY, P.C. 2018 Basic physical processes and reduced models for plasma detachment. *Plasma Phys. Control. Fusion* **60** (4), 044022.
- STANGEBY, P.C. 2000 *The Plasma Boundary of Magnetic Fusion Devices*, vol. 224. Institute of Physics.
- VAN DER MAATEN, L. & HINTON, G. 2008 Visualizing data using t-SNE. *J. Machine Learning Res.* **9** (11).
- WANG, L., XU, X.Q., ZHU, B., MA, C. & LEI, Y.-A. 2020 Deep learning surrogate model for kinetic Landau-fluid closure with collision. *AIP Adv.* **10** (7), 075108.
- WIESEN, S., REITER, D., KOTOV, V., BAELMANS, M., DEKEYSER, W., KUKUSHKIN, A.S., LISGO, S.W., PITTS, R.A., ROZHANSKY, V., SAIBENE, G., *et al.* 2015 The new SOLPS-ITER code package. *J. Nucl. Mater.* **463**, 480–484.
- WISING, F., KNOLL, D.A., KRASHENINNIKOV, S.I., ROGNLIEN, T.D. & SIGMAR, D.J. 1996 Simulation of detachment in ITER-geometry using the UEDGE code and a fluid neutral model. *Contrib. Plasm. Phys.* **36** (2–3), 309–313.
- WISING, F., KRASHENINNIKOV, S.I., SIGMAR, D.J., KNOLL, D.A., ROGNLIEN, T.D., LABOMBARD, B., LIPSCHULTZ, B. & MCCracken, G. 1997 Simulation of plasma flux detachment in Alcator C-Mod and ITER. *J. Nucl. Mater.* **241**, 273–277.
- ZHAO, M., ROGNLIEN, T., ZHU, B., MEYER, W., XU, X., DUDSON, B. & LI, N. 2022 Sub-supersonic transitions in the tokamak scrape-off layer and bifurcation solutions. *Phys. Plasmas*, in preparation.