J. Plasma Phys. (2022), *vol*. 88, 895880604 © The Author(s), 2022.

Published by Cambridge University Press

doi:10.1017/S0022377822001192

Data-driven model discovery for plasma turbulence modelling

I. Abramovic[1,2,†], **E.P. Alves**[3] **and M. Greenwald**[2]

[1]University of Technology Eindhoven, De Zaale, 5612 AJ Eindhoven, The Netherlands

[2]Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA

[3]Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA

(Received 31 May 2022; revised 15 October 2022; accepted 8 November 2022)

An important problem in nuclear fusion plasmas is the prediction and control of turbulence which drives the cross-field transport, thus leading to energy loss from the system and deteriorating confinement. Turbulence, being a highly nonlinear and multiscale process, is challenging to theoretically describe and computationally model. Most advanced computational models fall into one of the two categories: fluid or gyro-kinetic. They both come at a high computational cost and cannot be applied for routine simulation of plasma discharge evolution and control. Development of reduced models based on (physics informed) artificial neural networks could potentially fulfil the need for affordable simulations of plasma turbulence. However, the training requires an extensive data base and the obtained models lack extrapolation capability to scenarios not originally encountered during training. This leads to reduced models of limited validity which may not prove adequate for predicting scenarios in future machines. In contrast, we explore a data-driven model discovery approach based on sparse regression to infer governing nonlinear partial differential equations directly from the data. Our input data are generated by simulations of drift-wave turbulence according to the Hasegawa–Wakatani and modified Hasegawa–Wakatani models. Balancing model accuracy and complexity enables the reconstruction of the systems of partial differential equations accurately describing the dynamics simulated in the input data sets. Sparse regression is not data hungry and can be extrapolated to unexplored parameter ranges. We explore and demonstrate the potential of this approach for fusion plasma turbulence modelling. The findings show that the methodology is promising for the development of reduced and computationally efficient turbulence models as well as for existing model cross-validation.

Key words: fusion plasma, plasma nonlinear phenomena, plasma simulation

1. Introduction

Creating an energy source by means of harnessing energy released in nuclear fusion reactions has been a long outstanding goal in energy research. The prospect of clean

† Email address for correspondence: iabramovic@psfc.mit.edu

energy production with virtually inexhaustible fuel reserves has motivated the efforts to overcome the many engineering challenges and to fill the gaps in our understanding of relevant fundamental processes.

For practical fusion applications hydrogen isotopes – deuterium and tritium – need to be heated to extreme temperatures at which the fuel becomes a high-temperature plasma. The plasma can be thermally insulated and confined by a strong magnetic field, which is why the most advanced fusion concepts rely on magnetic confinement. The most promising is the tokamak: a device with a toroidal geometry featuring closed magnetic flux surfaces traced out by the field lines. The tokamak field geometry reduces the loss of charged particles in the direction perpendicular to the confining field. The experiments have shown that the quality of plasma confinement scales with the confining field strength and that it is determined by mechanisms which drive cross-field transport from the core to the edge and, in this way, drive the energy loss from the system. Thus our capability of putting fusion energy on the grid has been limited by both our capability to build large-scale high-field fusion magnets and our understanding of cross-field transport mechanisms – in particular of turbulence. Massachusetts Institute of Technology's SPARC (Creely *et al.* 2020) tokamak currently under development will incorporate high-temperature superconductors (HTS) (Hartwig *et al.* 2020) into large-scale high-field fusion magnets[1] and will realize a tokamak magnetic field strength over a factor of 2 larger than presently possible. The higher magnetic field will improve confinement and overall performance but it will also put the reactor into an unexplored parameter range when it comes to turbulence in the plasma edge. For this reason, edge turbulence modelling and its reliable extrapolation to SPARC conditions is critical.

Nonlinear gyrokinetic models (Wang *et al.* 2006; Brizard & Hahm 2007; Garbet *et al.* 2010; Cheng *et al.* 2020) capture the microscopic plasma behaviour and can simulate the plasma dynamics but at a high computational cost. Computationally less expensive fluid models, in certain parameter ranges of interest, can be applied for modelling the large-scale plasma dynamics but at the cost of excluding potentially important microscopic processes. The development of new reduced plasma models that optimally balance between physical accuracy and computational complexity are therefore essential for modelling these systems. However, their theoretical development is hindered by the complexity and the nonlinearity of plasma turbulence where first principle derivations are difficult and often elusive.

Machine learning and data science tools are offering a new approach to develop these models based on the data. Valuable results for fusion plasmas have been achieved by using neural networks trained on extensive data bases either generated by kinetic codes (van de Plassche *et al.* 2020) or experimental data (Churchill *et al.* 2020). Such neural networks are fast and for this reason useful in practical applications such as routine simulation of tokamak discharge evolution and control. Limitations of the neural network approach are its necessity for an extensive training data base and the lack of interpretability of the learning algorithms. The extrapolation of the results to scenarios not originally present in the training data base present a challenge although significant efforts are being made to improve the extrapolation capabilities of neural networks (Raissi, Perdikaris & Karniadakis 2019; Moseley, Markham & Nissen-Meyer 2020; Moseley, Nissen-Meyer & Markham 2020). For these reasons the application of neural networks to complex systems modelling leads to reduced models of limited validity. Other data-driven methods exist which could potentially be applied to turbulence in fusion plasmas. Among others

[1]Magnets developed for SPARC will be the first ever large-scale HTS magnets useful also in various applications other than fusion. For more details on SPARC see Creely *et al.* (2020).

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

these include: nonlinear Laplacian spectral analysis (Giannakis & Majda 2012), Koopman analysis (Mezic 2005, 2013), dynamic mode decomposition (Rowley *et al.* 2009; Schmid 2010; Kutz *et al.* 2016), symbolic regression (Bongard & Lipson 2007; Schmidt & Lipson 2009; Schmidt *et al.* 2011; Daniels & Nemenman 2015*a*,2015*b*) and sparse regression (Brunton, Proctor & Kutz 2016; Rudy *et al.* 2017).

In this paper we explore a novel data-driven approach for the discovery of reduced plasma models based on sparse regression (Brunton *et al.* 2016; Rudy *et al.* 2017). Our approach also stems from machine learning methods and attempts to infer nonlinear differential equations from synthetic time-series data. Sparse regression in theory selects parsimonious models by balancing model accuracy and complexity, enabling identification of governing partial differential equations directly from the data. The identified equation terms are dominant features selected from a matrix which spans the feature space containing all equation candidate terms. Unlike the artificial neural network or other data-driven model discovery methods (DDMD), sparse regression does not require large amounts of data. Moreover, the output of this algorithm is an interpretable partial differential equation, which offers insight into the underlying physics and has a greater capability to generalize beyond the training data. In addition, the integral formulation has more recently been proposed to offer robustness to noise (Schaeffer & McCalla 2017; Alves & Fiuza 2020; Reinbold, Gurevich & Grigoriev 2020; Messenger & Bortz 2021) crucial for applications to experimental data.

Sparse regression has been shown to successfully recover the fundamental hierarchy of plasma physics models from the kinetic Vlasov equation to single-fluid magneto-hydrodynamics from first principles (Alves & Fiuza 2020; Kaptanoglu *et al.* 2021). Here, we extend that work by applying it to resistive drift-wave turbulence simulations based on the Hasegawa–Wakatani (HW) and modified Hasegawa–Wakatani (mHW) models (Hasegawa & Mima 1978; Wakatani & Hasegawa 1984; Hasegawa & Wakatani 1987). These models couple the continuity and vorticity equations in reduced magnetohydrodynamics to describe drift-wave turbulence near the edge of a tokamak plasma. They are important yet simplified models and are therefore often used for benchmarking of fluid based turbulence simulations. The successful application of DDMD to HW and mHW shows the potential of this approach for inferring reduced but accurate descriptions of plasma turbulence which are directly relatable to analytic theory. We use the data generated by numerically solving the HW and mHW equations as input for our DDMD procedure and demonstrate that we can recover the correct set of coupled partial differential equations. To further probe the potential for applications to experimental data, we repeat the procedure at increasing noise levels and lay the groundwork for the development of feature space completeness analysis. Detailed description of our method as well as the HW turbulence model is given in § 2. In § 3 we present our main findings and explore the issue of feature space completeness. The discussion and conclusions are given in §§ 4 and 5.

2. Method

2.1. *DDMD and sparse regression*

In general, sparse (symbolic) regression is a type of machine learning algorithm which searches a predefined feature space in order to output mathematical expressions which best describe the input dataset. The objective is to use sparse regression to infer the underlying partial differential equations (PDEs) from the data. The inference of a PDE is done by regression on a predefined line up of candidate PDE terms – the feature space. The method used for this work is suited to be applied to time-series data with spatio-temporal variation,

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

in particular data sets containing the evolution of the system dynamics in the nonlinear regime. The feature space Θ is defined as a matrix in which each column represents a scaled feature: a numerical estimate of a variable, its derivatives, linear combinations of variables and derivatives or polynomial combinations of terms up to a specified order, over a data sample over time. The algorithm discovers both the model structure and the model parameters. In practice prior knowledge and physical intuition about the systems dynamics informs the design of candidate terms – features – included in the feature space. However, in theory, prior knowledge about the system dynamics is not used. This makes the approach potentially suitable for both model validation as well as probing for new underlying physics.

The balance between the discovered model accuracy and complexity is achieved via Pareto analysis that prevents overfitting. In practice this leads to the optimization favouring models with less selected features (equation terms) over the ones with more features at the same accuracy. The cost function being minimized is of the following general form:

$$||\partial_t \langle F \rangle - \langle \Theta \rangle \xi||_2^2 + \lambda ||\xi||_0, \tag{2.1}$$

where the variable F is a quantity whose dynamics is governed by a PDE that we seek to discover. The terms of this equation are the features in Θ selected by multiplication with the sparsest vector ξ which contains the model parameters (coefficients in front of the corresponding terms in the PDE). The vector elements of ξ are computed using a thresholded least-squares algorithm (Brunton *et al.* 2016), while its level of sparsity is obtained by means of a 30-fold cross-validation procedure. Vector λ contains 100 regularization parameters used in sequentially thresholded least-squares optimization of ξ. The values of the regularization parameters are logarithmically increased from 10^{-6} up to 1 to obtain increasingly sparser solutions of ξ. Each candidate feature in Θ and the time derivatives vector $\partial_t F$ are approximated by central finite differencing and integrated over volumes on the sampled data. The angle brackets denote averaging over the sampled volumes. Therefore the governing PDE is discovered in its integral form, this methodology has been described in detail in Alves & Fiuza (2020) where it is applied to a hierarchy of basic plasma physics models.

2.2. *The HW model*

In our particular case the input data are generated through direct numerical simulation of resistive drift-wave turbulence according to two-dimensional (2-D) HW and mHW models (Hasegawa & Mima 1978; Wakatani & Hasegawa 1984; Hasegawa & Wakatani 1987) in the (x, y) plane perpendicular to the magnetic field along the z axis. These models are a good test bed for our model discovery method for two main reasons. Firstly, these are relevant models providing a description of drift-wave turbulence near the tokamak plasma edge. Secondly, they are simple consisting of two coupled PDEs with a relatively small number of terms and therefore allow straightforward interpretation and checking of obtained results. The HW couples the continuity and vorticity equations in reduced MHD (Wakatani & Hasegawa 1984; Hasegawa & Wakatani 1987; Biskamp & Zeiler 1995)

$$\frac{\partial n}{\partial t} = -[\phi, n] + \alpha(\phi - n) - \kappa \frac{\partial \phi}{\partial y} - D_n \nabla^4 n, \tag{2.2}$$

$$\frac{\partial \omega}{\partial t} = -[\phi, \omega] + \alpha(\phi - n) - D_\omega \nabla^4 \omega, \tag{2.3}$$

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

where n and ω represent the electron density and vorticity,[2] respectively, ϕ stands for the electrostatic potential and the constants α and κ are the electron adiabaticity and the background density gradient driver. The Poisson bracket for two functions of phase space and time, f and g, is defined in the usual manner $[f, g] = \hat{z} \times \nabla f \cdot \nabla g$. Similarly, mHW couples the continuity and vorticity equations but captures the electron adiabatic response with the substitutions $n \to \tilde{n}$ and $\omega \to \tilde{\omega}$ where the new variables are defined by

$$\tilde{f} = f - \frac{1}{L_y} \int f \, \mathrm{d}y \tag{2.4}$$

and the equations being numerically solved become

$$\frac{\partial n}{\partial t} = -[\phi, n] + \alpha(\tilde{\phi} - \tilde{n}) - \kappa \frac{\partial \phi}{\partial y} - D_n \nabla^4 n, \tag{2.5}$$

$$\frac{\partial \omega}{\partial t} = -[\phi, \omega] + \alpha(\tilde{\phi} - \tilde{n}) - D_\omega \nabla^4 \omega. \tag{2.6}$$

Inclusion of the electron adiabatic response enables (for certain parameter values) the simulation of elongated asymmetric vortex modes, so-called, zonal flows.[3] The introduced substitutions as defined by (2.4) relaxes the coupling between the electric potential ϕ and the density n for these modes as they are not subject to the coupling arising from the parallel momentum equation. Zonal flows are driven by nonlinear energy transfer from drift waves and represent a self-regulation agent for drift-wave transport and turbulence (Diamond *et al.* 2005). Zonal flows play an important role in fusion plasmas because they suppress plasma turbulence and thus reduce the transport driven by it in turn improving plasma confinement.

It should be noted here that the standard drift-wave normalizations (Wakatani & Hasegawa 1984; Biskamp & Zeiler 1995) were used: $\phi \longrightarrow (e\phi/T_e)L_n/\rho_s$,$n \longrightarrow (n/n_0)L_n/\rho_s$, $t \longrightarrow tc_s/L_n$, $\nabla_\perp \longrightarrow \rho_s \nabla_\perp$ and $\nabla_\parallel \longrightarrow L_\parallel \nabla_\parallel$, where ρ_s is the ion Larmor radius, n_0 the background electron density, e the electron charge, c_s the ion acoustic velocity, ν_{ei} the electron ion collision rate and the L_n the scale length of the density background. The parallel scale length is defined by $L_\parallel = (L_n T_e/m_e c_s \nu_{ei})^{1/2}$.

2.3. *Simulation set-up*

The simulation outputs the changes in plasma density, vorticity and potential in the (x, y) plane. The size of the simulated domain was $16\pi \times 16\pi$. The temporal resolution was varied while keeping the total simulated time t constant and sufficient to allow for nonlinear effects to develop and evolve. Once the input data sets are available what follows is setting up the data sampling routine. In our particular case the data were sampled with varying sampling density both spatially (distribution and number of sampling volumes) and temporally (time-series length). Uniform sampling was applied only in space while the time sampling was non-uniform/random within the specified time interval. This time sampling allows enables capturing both fast and long-time scale features. The sampling is a routine which may require adjustment tailored to the input data structure and the expected underlying physics. For some problems the uniform sampling might not be appropriate and one might choose to use threshold-based sampling or targeted sampling restricted to predefined regions of the phase space.

[2]Note here that the vorticity is given by $\omega = \nabla_\perp^2 \phi$ and that one should not dismiss the significance of the background magnetic filed in the HW model (Wakatani & Hasegawa 1984; Hasegawa & Wakatani 1987; Biskamp & Zeiler 1995).

[3]Although it should be noted that zonal flows can in principle be obtained with the standard HW model at low α values.

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

The features (the potential terms in the sought after PDE) need to be chosen and evaluated on the sampled data points. The choice of features to include in the Θ matrix relies on the physical intuition about the problem at hand. Once the feature matrix is constructed, sequentially thresholded least-squares regression is applied to the sampled data in order to obtain the sparsity vector ξ which reveals the retained features and the numerical values of their corresponding coefficients (for code structure diagram see figure 1). Plotting the fraction of variance unexplained (FVU) as a function of the number of retained features reveals the discovered PDE. The FVU is in this work defined as the ratio of the model's mean squared error and the variance of the partial time derivative of the dependent variable (density n or vorticity ω). Therefore, for the density equation we have the following definition of FVU:

$$\mathrm{FVU} = \frac{\mathrm{MSE(model)}}{\mathrm{var}\left(\dfrac{\partial n}{\partial t}\right)}, \tag{2.7}$$

and similarly for the vorticity we have

$$\mathrm{FVU} = \frac{\mathrm{MSE(model)}}{\mathrm{var}\left(\dfrac{\partial \omega}{\partial t}\right)}. \tag{2.8}$$

A significant reduction in FVU is observed for the optimal number of features necessary for describing the dynamics underlying the input data.

2.4. *Structural similarity index*

Successful regression requires that the sampled data set contains sufficient change in the system dynamics in the nonlinear regime. To quantify this change in our sample we use the so-called, structural similarity index (SSI). The index was originally defined for the purpose of image quality assessment based on degradation of structural information (Wang *et al.* 2004). It is calculated for two given images and has the value in the range $[1, -1]$ where the extremes of the range indicate that the two images are identical (SSI $= 1$) or completely different (opposite) (SSI $= -1$). The SSI for two images is defined by the following expression:

$$\mathrm{SSI}(\boldsymbol{x}, \boldsymbol{y}) = [l(\boldsymbol{x}, \boldsymbol{y})]^{\alpha} \cdot [c(\boldsymbol{x}, \boldsymbol{y})]^{\beta} \cdot [s(\boldsymbol{x}, \boldsymbol{y})]^{\gamma}, \tag{2.9}$$

where x and y are two images of size $N \times N$ which are being compared based on the luminance l, contrast c and structure s functions weighed by the exponents α, β and γ. The luminance function $l(\boldsymbol{x}, \boldsymbol{y})$ is given by

$$l(\boldsymbol{x}, \boldsymbol{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{2.10}$$

where μ denotes the average of all pixel values (mean luminance intensity) and the constant C_1 prevents instability for denominators approaching 0. Similarly, the contrast

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

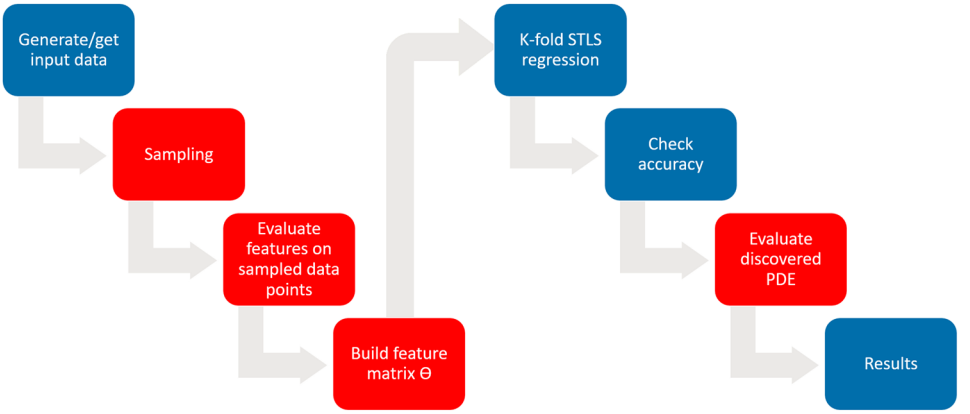

FIGURE 1. Code structure: the arrows denote the flow of information and the blocks depict the sequence of steps performed by the sparse regression (SR) algorithm. The colour coding refers to parts of the code which might need to be adapted to a particular physics problem depending on the input data structure and content (red) and the parts which essentially remain the same regardless of the input data structure and content (blue).

function $c(\boldsymbol{x}, \boldsymbol{y})$ is given by

$$c(\boldsymbol{x}, \boldsymbol{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{2.11}$$

where σ denotes the standard deviation of all pixel values and C_2 is a constant. Finally, the structure function $s(\boldsymbol{x}, \boldsymbol{y})$ is given by

$$s(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{2.12}$$

where the correlation coefficient σ_{xy} is equal to

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) \tag{2.13}$$

Equations (2.9)–(2.13) are given here for the sake of completeness, for more details on the derivation and the SSI metric we refer the reader to the original reference (Wang *et al.* 2004). Our data is a series of snapshots of the configuration space which can be converted to a series of images (see figure 2) to which a sequence of SSI can be attributed.

3. Results

3.1. *Data generation and sampling*

Simulation data were obtained by running the phw.f90 code[4] on the MIT Engaging cluster. The code solves (2.2) and (2.3) in two dimensions for the density and vorticity. The simulated domain was temporally resolved with $\Delta t = 0.01$ ms and spatially resolved with $\Delta x = 0.19$ and $\Delta y = 0.19$ on a grid of size 256×256. The total simulated time allowing

[4]The phw.f90 is a reduced version of the Global Drift Ballooning code (Fisher & Rogers 2017; Zhu, Francisquez & Rogers 2017) and has been benchmarked against BOUT++ and MFEM.

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

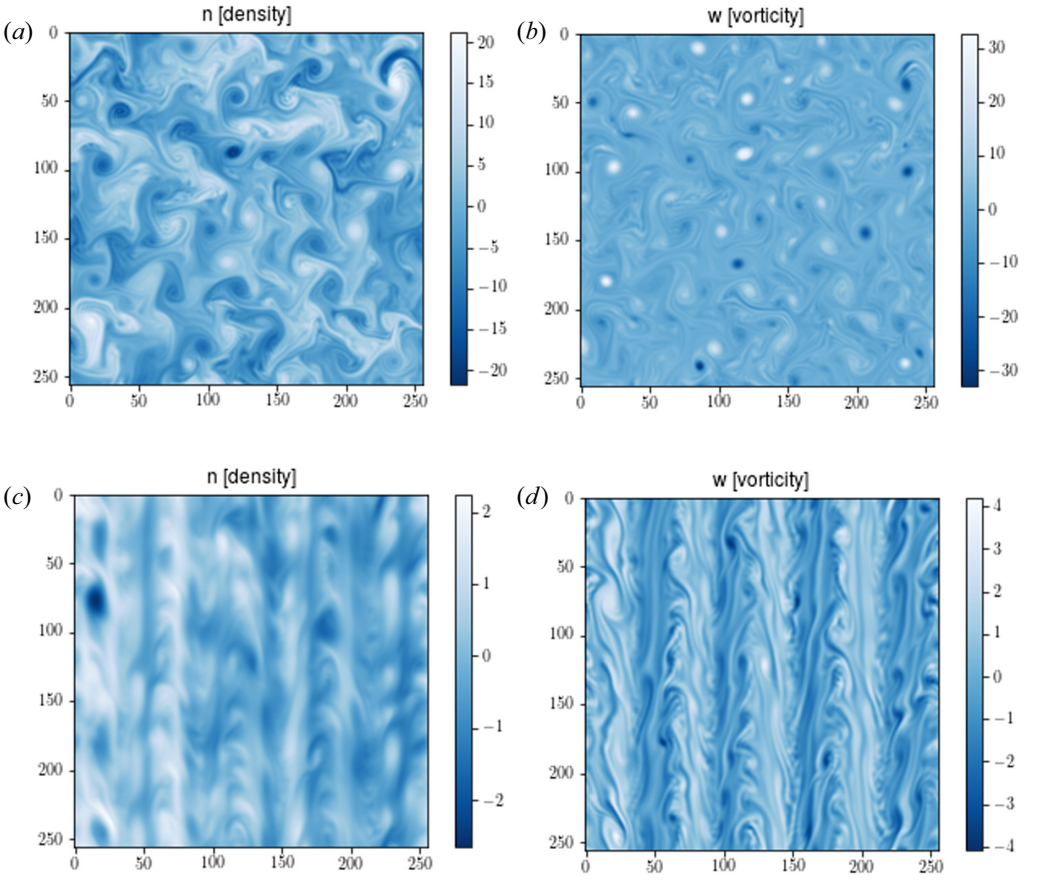

FIGURE 2. The HW (*a*,*b*) and mHW (*c*,*d*) density and vorticity data.

Resolution	α	κ	D_n	D_ω	dt
256×256	0.1	1	10^{-4}	10^{-4}	0.01

TABLE 1. Simulation set-up and correct model coefficients.

for the nonlinear behaviour to develop was $t = 1000$. Snapshots (time frames) of the density, potential and vorticity were collected and stacked together forming a long time series containing the evolution of the system over the total simulated time. Values of other input parameters used for the simulation data in this work are given in table 1. Prior to building of the feature matrix Θ, the dynamics captured by the data has to be sampled adequately in phase space and in time. Designing the sampling routine is done with the input data structure in mind. In this work the sampling is volumetric and uniform: volumes sampled in phase space are uniformly distributed over the simulation grid. Both the sample grid size and the length of the time series from which the data points are sampled were varied (see figure 4). The results of this sensitivity study are given in § 3.4.

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

3.2. *Constructing the feature space* Θ

Once the data sampling is complete, we proceed to construct the matrix Θ containing the possible terms in the sought after PDE. Each column of the matrix represents a feature evaluated at each sampled data point. For our HW data, the primary features are the variables density n, electrostatic potential ϕ and vorticity ω. Secondary features we considered are the spatial gradients of the primary features up to fourth order and the polynomial combinations of the primary and secondary features up to n order. In total we considered 120 features (columns of Θ) in our analysis. The second-order centred finite differencing is applied to sampled data neighbouring points and is used to evaluate the derivative terms. A 30-fold sequentially thresholded least-squares regression is applied to the sampled data divided into training, test and validation sets in order to obtain the sparsity vector.

3.3. *Results of sparse regression*

The 30-fold sequentially thresholded least-squares regression applied to our Θ feature matrix yields the accuracy/complexity curves and the sparsest vector ξ. The accuracy/complexity curves show the dependence of FVU (accuracy) on the number of features retained (complexity), while the multiplication of Θ and ξ, for given complexity or number of features retained, yields the features with their corresponding coefficients. The last step enables the reconstruction of the PDE system of (2.2) and (2.3). The Θ matrix contained all features that were necessary to completely describe the dynamics in the input data. The success or failure of the procedure is indicated by the inflection in the accuracy/complexity curves, for the HW data, presented in figure 4. The FVU suddenly drops due to the thresholding of a dominant term, this marks the trade-off between model accuracy and complexity. The pronounced inflection in the accuracy/complexity curves is observed at the model complexity of 5 terms for the density equation and at the model complexity of 4 terms for the vorticity equation (see figure 3 and tables 2 and 3). This corresponds to the complete system of HW equations (2.2) and (2.3) given here in their explicit form for completeness

$$\frac{\partial n}{\partial t} = \frac{\partial \phi}{\partial y}\frac{\partial n}{\partial x} - \frac{\partial \phi}{\partial x}\frac{\partial n}{\partial y} + \alpha\phi - \alpha n - \kappa\frac{\partial \phi}{\partial y} - D_n\nabla^4 n, \tag{3.1}$$

$$\frac{\partial \omega}{\partial t} = \frac{\partial \omega}{\partial y}\frac{\partial n}{\partial x} - \frac{\partial \omega}{\partial x}\frac{\partial n}{\partial y} + \alpha\phi - \alpha n - D_\omega\nabla^4\omega. \tag{3.2}$$

3.4. *Sensitivity study*

Once the regression has been shown to successfully recover the complete system of HW equations, a sensitivity study has been designed to test the method's robustness to three main aspects: spatial sampling domain, length of time series considered in the sampling and the noise level. Each of the aspects has been studied independently from the others; for each a reference regression set-up has been fixed and only the parameter directly related to the aspect studied has been varied.

3.4.1. *Spatial domain size*

Assessing the sensitivity to the spatial sampling domain size entailed reducing the domain size from the optimal, which covered most of the grid, to a size comparable to characteristic lengths in the data (such as the eddy dimensions) (see figure 4). An increase in the error on the model coefficients as well as a change in model-form variance (change in the sparsity vector ξ) was observed when reducing the spatial sampling domain

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

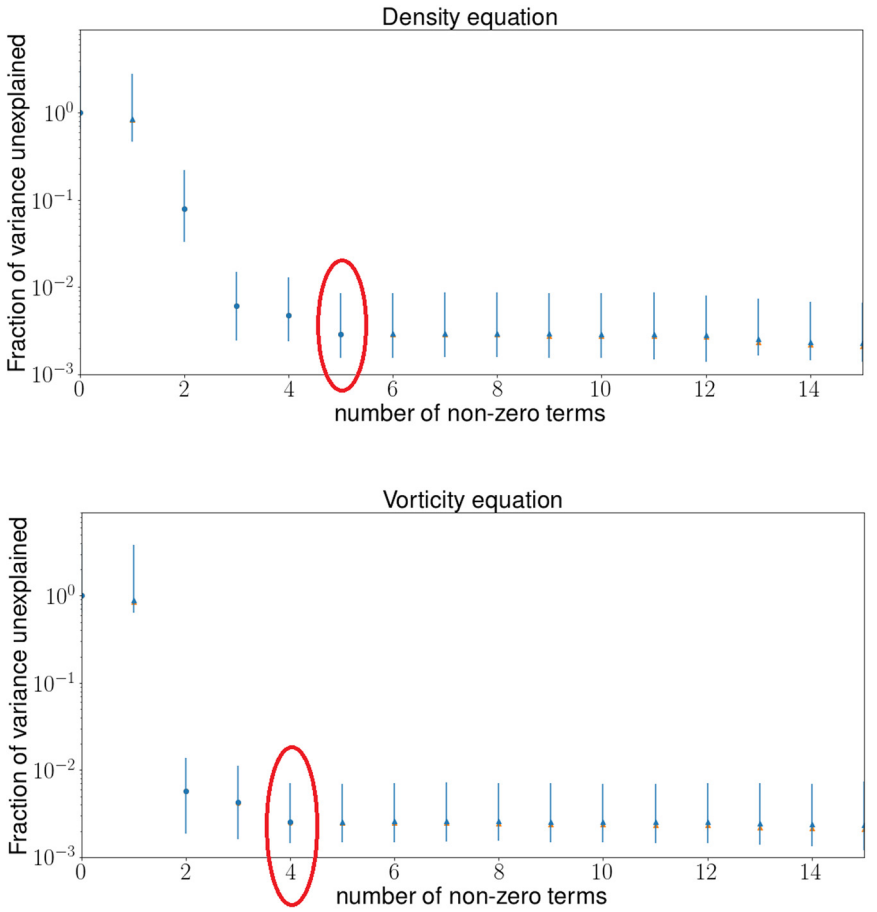

FIGURE 3. Main result on the noise-less synthetic data. The two different markers, circles and triangles, represent model-form variance obtained in the cross-validation procedure (variance in the sparsity vector ξ). The red ellipses circle the points of optimal accuracy/complexity trade-off and correspond to identification of correct features of the HW system of PDE. The 'error' bars represent the minimum and maximum FVU encountered during cross-validation.

Sampling grid	# time frames	Recovered features	Avg. coef. error	K-fold cross val.
241×241	700	5/5	$\approx 2\,\%$	30

TABLE 2. Results of SR for the density equation of the HW and mHW models.

Sampling grid	# time frames	Recovered features	Avg. coef. error	K-fold cross val.
241×241	700	4/4	$\approx 2\,\%$	30

TABLE 3. Results of SR for the vorticity equation of the HW and mHW models.

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

Spatial sampling domain size	Total recovered features	Avg. recovered coef. error
241×241	9/9	$\approx 2\,\%$
206×206	9/9	$\approx 2\,\%$
181×181	7/9	$\approx 3\,\%$
156×156	6/9	$\approx 5\,\%$
50×50	5/9	$\approx 6\,\%$

TABLE 4. Results of the sensitivity to the spatial domain size of the sampling region.

Total time-series length	Sampled time-series length	Total recovered features	Avg. recovered coef. error
4000	100	7/9	$\approx 5\,\%$
4000	200	9/9	$\approx 6\,\%$
4000	500	9/9	$\approx 4\,\%$
4000	700	9/9	$\approx 5\,\%$
4000	1500	7/9	$\approx 4\,\%$
4000	2000	5/9	$\approx 3\,\%$

TABLE 5. Results of the sensitivity to the sample time-series length.

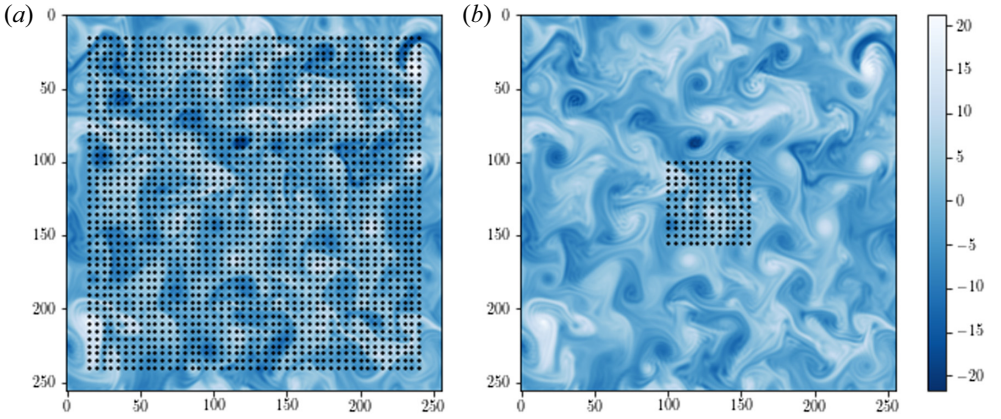

FIGURE 4. Sensitivity to spatial sampling domain size: largest (*a*) and smallest (*b*) sampling regions used in the study.

size (see table 4). Here, the number of sampled points per sampling volume has not been varied in order to probe solely the effect of sampling domain size reduction. However, the results can be improved by increasing the number of sampled points per sampling volume until the mean coefficient error saturates (provided that the sampling domain size remains comparable to the characteristic lengths in the data). This would be analogue to the increase in number of points per volume for the case of clean data (see table 6). More details on this technique have been given in the supplementary material of Alves & Fiuza (2020).

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

	Volume					
p	$5 \times 5 \times 5$	$6 \times 6 \times 6$	$7 \times 7 \times 7$	$8 \times 8 \times 8$	$9 \times 9 \times 9$	$10 \times 10 \times 10$
0 %	≈3 %	≈2 %	≈1 %	≈1 %	≈1 %	≈1 %
6 %	≈5 %	≈3 %	≈2 %	≈3 %	≈2 %	≈2 %
12 %	≈10 %	≈8 %	≈5 %	≈4 %	≈4 %	≈4 %
18 %	≈14 %	≈12 %	≈9 %	≈9 %	≈9 %	≈5 %

TABLE 6. Average error on the recovered feature coefficients from noisy data.

3.4.2. *Time-series length*

Assessing the sensitivity to the sampled time-series length entailed gradually increasing the number of frames to sample from (see table 5) and establishing the minimal number necessary for recovery of the HW system of equations. This minimal length of the sampled time series is determined by the change or evolution of the system dynamics contained in it. The quantification of this change is not trivial and to our knowledge has not been examined in available literature on application of sparse regression to identification of PDE. However, it is important for establishing whether a sample is suitable for sparse regression and it is useful in reducing the sample size. The metric proposed and used in this work is the rate of change in the slope of the, previously introduced, SSIs calculated over samples of differing length. Namely, we set the reference image for SSI calculation to be the first frame in our data set. We then chose the sampling time-series length (the number of frames) and compare our reference frame with consecutive frames from the series by calculating the SSI. The sequence of SSI indices obtained characterizes the system evolution from the reference frame to the last frame in the sample. As expected the value of SSI will decrease and level off as the similarity between the reference frame and the consecutive frames becomes lesser due to the system evolving further from its initial state. For the sparse regression to successfully recover the HW system of equations, it proves optimal to set the length of the sample to the length for which the SSI value levels off. To find this value we look at the extremum in rates of change in the slope of the SSI – the first derivative of SSI for samples of different length. The SSI levels of when the first derivative of SSI becomes ≈0 see figure 5.

3.4.3. *Noise sensitivity*

A particularly interesting application of sparse regression would be application directly to experimental data. On the one hand, due to the possibility of discovering reduced turbulence models useful in practical applications such as control, and on the other due to the possibility of discovering new physics. However, experimental data always contain noise and it is therefore necessary to investigate to what extent the noise corrupts the recovery of underlying PDEs. To do this, prior to applying the sparse regression, we added Gaussian noise to our simulation data. The noise had a mean $\mu = 0$ and a standard deviation defined as a function of the input data: $\sigma = \sigma(\text{input data}) \cdot p$, where p denotes percentage. The noise level was increased from 0 % up to 18 % while the sampling settings were kept fixed. As the noise level is increased the model error increases and the inflection in the accuracy/complexity curve is reduced until it vanishes completely, as shown in figure 6. With noise, model-form variance and the error of the model coefficients increase while the number of recovered features progressively decreases. The model-form variance is observed not to be affected by the size of the integration volumes. However, the model

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

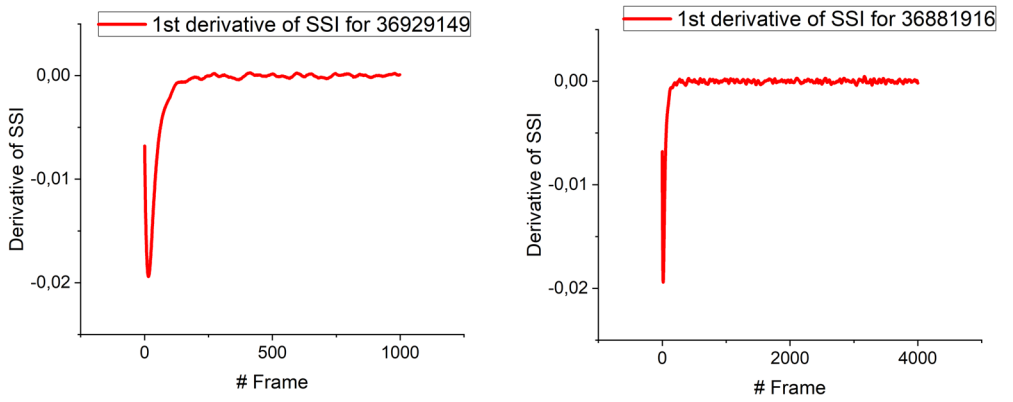

FIGURE 5. First derivative of the SSI over the sample. The SSI is calculated between the first data frame and every consecutive data frame throughout the sample. The curve indicates the evolution of the dynamics in the data starting from the first frame.

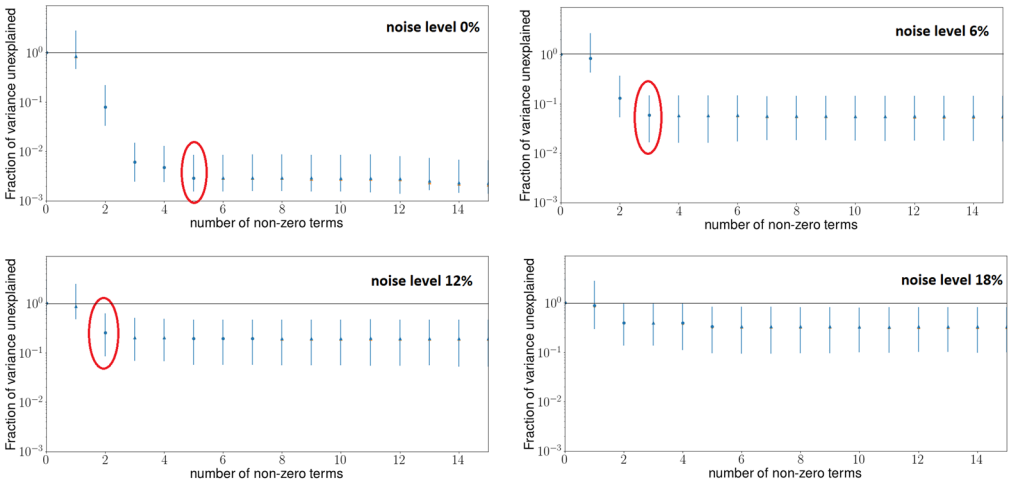

FIGURE 6. The FVU for the density equation of the HW model for noisy data. The level of noise is increased from 0 % to 18 % (see table 6 for details). The inflection in the accuracy/complexity curve is reduced as the noise level is increased. The number of recovered terms of the density PDE progressively drops from 5 to 2.

coefficient errors are affected by the changes in integration volumes and can be reduced up to a certain extent, but not indefinitely, by increasing their size.

It is important to note that, in the presence of noise, fewer terms from the governing equations may be recovered. This is illustrated in table 7 for 6 % noise added to the input data set. As the size of the sampling volumes is increased the algorithm is able to recover progressively more terms. For the example of a relatively low noise level this eventually leads to the recovery of all features and accompanying coefficients. The dominant features recovered in the example for the smallest sampling volume considered ($5 \times 5 \times 5$) are: $(\partial\phi/\partial y)(\partial n/\partial x)$, $(\partial\phi/\partial x)(\partial n/\partial y)$, $\partial\phi/\partial y$, $(\partial\omega/\partial y)(\partial n/\partial x)$ and $(\partial\omega/\partial x)(\partial n/\partial y)$.

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

	Volume			
p	$5\times5\times5$	$6\times6\times6$	$7\times7\times7$	$8\times8\times8$
6 %	5/9	7/9	7/9	9/9

TABLE 7. Number of features recovered in the presence of 6 % noise and increasing sampling volume size.

3.5. *Feature space completeness*

This section has been added for the sake of completeness as a comment specifically on the model bias error – associated with missing features in the Θ matrix (Alves & Fiuza 2020). Spatio-temporal distribution of the model bias error is defined as

$$\epsilon_{n_m} = \left|\left|\frac{\mathrm{d}n}{\mathrm{d}t} - \Theta\xi_{n_m}\right|\right|_2^2, \quad \epsilon_{\omega_m} = \left|\left|\frac{\mathrm{d}w}{\mathrm{d}t} - \Theta\xi_{w_m}\right|\right|_2^2. \tag{3.3}$$

If Θ is complete, contains all features needed to describe the input data set, the model bias error vanishes for the correct model

$$\lim_{m\to5} ||\xi_{n_m}||_2^2 = \eta_n^2, \quad \lim_{m\to4} ||\xi_{w_m}||_2^2 = \eta_\omega^2, \tag{3.4}$$

where η_n and η_ω represent intrinsic noise.

4. Discussion and outlook

We have shown that the sparse regression algorithm described in Alves & Fiuza (2020) can be applied to synthetic plasma drift-wave turbulence data to extract the governing PDEs. Two parts of the algorithm are particularly important for successful PDE identification: the sampling and the construction of the Θ feature matrix. Sampling used in this work is uniform across the simulation grid meaning that the sampling volumes are uniformly distributed (as depicted in figure 4), each volume containing a number of data points in phase space and time.[5] Sensitivity to both spatial and temporal sampling region bounds has been studied. For successful SR it suffices to set the spatial sampling domain size to be larger than the characteristic length. This, for turbulence data, is the typical eddy size. Similarly, the sampled time-series length can be related to the characteristic time in the data, in this case the time correlation length. In an attempt to define a systematic approach to choosing the sampling time-series length for successful PDE identification, we proposed the use of the SSI commonly found in applications involving image quality and data compression. This metric is applicable to synthetic turbulence 2-D data for quantification of the system evolution over time. Calculating the slope of SSI over the sampled time series enables the evaluation of a time series truncation threshold beyond which including more subsequent time frames does not improve the results of SR on the sample. Defining the spatio-temporal bounds for sampling in this way provides a methodology for minimizing the data requirements for successful SR on the sample. In the prospect of applying SR to synthetic data from more complex turbulence simulations, such as the full gyro-kinetic codes (Wang *et al.* 2006; Cheng *et al.* 2020) which produce large data sets in comparison with what has been used in this work, the SSI based threshold

[5]The cubic volumes are defined as $n\Delta x \times n\Delta y \times n\Delta t$, each volume containing n^3 points.

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

enables one to define the minimal part of a data set from which to sample in preparation for SR.

We recovered all the physical terms from the HW and mHW equations. Terms which were not recovered are the higher-order diffusion terms, namely $D_n\nabla^4 n$ and $D_\omega\nabla^4\omega$, which are terms added for numerical stability and have the purpose of dissipating the grid size eddies.[6] However, one can encounter data sets in which higher-order terms are physical and therefore relevant. We speculate that recovering such terms, a few orders of magnitude smaller than the leading terms, is possible and requires the use of targeted sampling: predominantly sampling from regions of phase space where fine effects are expected to be dominant.

The features in the Θ matrix were chosen based on domain knowledge and the matrix contained all the features[7] needed to fully describe the dynamics in the data. In cases when Θ does not contain all the dynamical terms for a particular problem, one should expand the feature library until the empirical signature[8] of a successful PDE identification is observed and the spatio-temporal error distribution is minimized. The spatio-temporal error distribution can further be used to asses the importance of discovered terms as described in § 3.5.

When the sampling and Θ construction are complete, SR can be applied to extract the governing equations in their integral form. The integral formulation strategy makes the approach suitable for application to noisy data as it mitigates the numerical differentiation errors through averaging over more sampled points. The HW and mHW models were successfully recovered with mean coefficient errors of $\approx$2 % for the clean data sets and 8 % for noisy data sets up to a significant level of added noise (up to 18 % noise) at the sampled volume size $5 \times 5 \times 5$. The increase of the size of sampled volumes reduces the effects of noise until the level of irreducible error is reached. At this level, as the noise is increased, SR recovers progressively fewer terms from the governing equations. Nevertheless, the retained terms correspond to the dominant processes underlying the dynamics. This is promising in the prospect of applying SR methodology directly to experimental data which inevitably contain noise. An important consideration in this context is the distinction between the measurements and variable noise. Presence of such noise could corrupt the discovered model form by selecting features from Θ characterizing the noise rather than solely the physical phenomenon of interest. One approach has been proposed in Fasel *et al.* (2022) and relies on statistical ensembling to improve robustness and accuracy for SR-based model discovery. In particular, the work relies on bootstrap aggregating to identify ensembles of ordinary and PDEs that govern the dynamics from noisy data. Inclusion probabilities for features of Θ can be obtained thus increasing the success rate for discovering the correct model and discarding the noise related terms. It remains to be investigated whether the approach improves on our integral formulation of SR for plasma turbulence simulation. In addition, future work will focus on applying the SR methodology to more complex turbulence simulations for development of computationally efficient reduced models. In the first instance, we will apply the methodology to the global drift ballooning model (Fisher & Rogers 2017; Zhu *et al.* 2017). This requires expanding the Θ matrix with rational function nonlinearities and their ratios, in order to increase the descriptive capacity of SR for the more complex simulated dynamics and

[6]The diffusion coefficients D_n and D_ω are three orders of magnitude smaller than the smallest recovered coefficient in the system of PDEs.

[7]It is worth emphasizing that the features are not mutually orthogonal functions but rather possible terms in the underlying differential equations.

[8]The empirical signature is defined as a pronounced inflection in the model accuracy/complexity curve signifying the optimal trade-off between model complexity and accuracy (Alves & Fiuza 2020; Fasel *et al.* 2022).

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

facilitate the application to experimental data. This application is of particular interest for control purposes. There are substantial data bases of experimental data of turbulent dynamics from fusion devices where the novel method could search for the underlying physical models. What has to be considered when applying this method to experimental data is that some coefficients are not constant over a wide range of plasma parameters. Therefore, when the method is applied to experimental data it can be used to recover the values of parameter-dependent coefficients. Based on prior knowledge about the data and data availability, we can distinguish among the following situations:

(i) We have spatio-temporal data available in a wide range of plasma parameters: in this case we intentionally retain a large number of degrees of freedom in Θ and apply the methodology to spatio-temporal data in different parameter ranges. This should enable to recover both the underlying physical model as well as the coefficients which depend on plasma parameters.
(ii) We have spatio-temporal data available in a narrow range of plasma parameters: in this case we need to proceed with caution and keep in mind that the coefficient can depend on the plasma parameters. The methodology in this case is primarily applied to recover the underlying dynamical terms but the interpretation of the recovered coefficients has to be made based on physical intuition and prior knowledge.

If we decide to assume a physics model describing the experimental data, this assumption can be used to reduce the number of degrees of freedom in our feature space Θ. The methodology can then be applied identify plasma parameters featuring in our assumed model.

Furthermore, we can use our methodology in an attempt to construct a reduced reconciling model for the plasma edge and core dynamics as this is still lacking due to the complexity of the collision operator and the asymptotic matching between fluid and gyrokinetic models. Furthermore, if we could use our method to point out what the most important terms in the feature space for turbulence data are, the result could inform reduced model selection from the myriad of already established reduced models. It should be pointed out that one can adopt a different approach to searching for the important terms in the feature space for turbulence data. Namely, valuable attempts have been made by using convolution neural networks such as the one described in Frezat *et al.* (2021).

5. Conclusion

In summary, we have presented a proof of principle study of the application of data-driven integral sparse regression to fusion plasma drift-wave turbulence simulations. We have shown that the methodology can successfully be applied to discover the system of PDEs governing the simulated plasma dynamics both from clean and noisy data. We have also conducted a sensitivity study of spatio-temporal characteristics of the sampling used for SR and established that the spatio-temporal sample bounds can be defined in relation to the minimal eddy size and correlation length characterizing the data. This approach ensures that the sampling phase-space region contains sufficient system evolution in the nonlinear regime for SR while at the same time ensuring one can use a subset of a large data set for the study. We have also discussed the completeness of the feature space Θ from which terms of the PDEs are selected by SR and formulated criteria for Θ completeness based on the spatio-temporal error distribution following the reasoning expressed in Alves & Fiuza (2020).

https://doi.org/10.1017/S0022377822001192 Published online by Cambridge University Press

Acknowledgements

The authors would like to thank Dr B. Zhou and Dr M. Francisquez for allowing their drift-wave turbulence codes to be used for producing input data sets for this study. As well as their support and suggestions which have improved this work and helped it reach its present form. Also a special thank you goes to Professor Dr N.L. Cardozo for his participation in discussions on the applicability and application of the presented methodology in fusion research. This publication is part of the project *Tackling Turbulence in the Innovative MIT Fusion Reactor Concept* (with project number 019.193EN.033 of the research programme (grant) *Rubicon* which is financed by the Dutch Research Council (NWO).

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

Declaration of interests

The authors report no conflict of interest.

REFERENCES

ALVES, E.P. & FIUZA, F. 2020 Data-driven discovery of reduced plasma physics models from fully-kinetic simulations. arXiv:2011.01927.

BISKAMP, D. & ZEILER, A. 1995 Nonlinear instability mechanism in 3D collisional drift-wave turbulence. *Phys. Rev. Lett*. **74** (5).

BONGARD, J. & LIPSON, H. 2007 Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **104** (24), 9943–9948.

BRIZARD, A.J. & HAHM, T.S. 2007 Foundations of nonlinear gyrokinetic theory. *Rev. Mod. Phys*. **79**.

BRUNTON, S.L., PROCTOR, J.L. & KUTZ, J.N. 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932.

CHENG, J., DOMINSKI, J., CHEN, Y., CHEN, H., MERLO, G, KU S.-H., HAGER, R., CHANG, C.-S., SUCHYTA, E., D'Azevedo, E., *et al.* 2020 Spatial core-edge coupling of the particle-in-cell gyrokinetic codes GEM and XGC. *Phys. Plasmas* **27**, 122510.

CHURCHILL, R.M., TOBIAS, B., ZHU, Y. & DIII-D TEAM 2020 Deep convolutional neural networks for multi-scale time-series classification and application to tokamak disruption prediction using raw, high temporal resolution diagnostic data. *Phys. Plasmas* **27**, 062510.

CREELY, A.J., GREENWALD, M.J., BALLINGER, S.B., BRUNNER, D., CANIK, J., DOODY, J., FÜLÖP, T., GARNIER, D.T., GRANETZ, R., GRAY, T.K., *et al.* 2020 Overview of the SPARC tokamak. *J. Plasma Phys*. **86**, 865860502.

DANIELS, B.C. & NEMENMAN, I. 2015*a* Automated adaptive inference of phenomenological dynamical models. *Nat. Commun*. **6** (1), 8133.

DANIELS, B.C. & NEMENMAN, I. 2015*b* Efficient inference of parsimonious phenomenological models of cellular dynamics using s-systems and alternating regression. *PLoS One* **10** (3), e0119821.

DIAMOND, P.H., ITOH, S.-I., ITOH, K. & HAHM, T. S. 2005 Zonal flows in plasma – a review. *Plasma Phys. Control. Fusion* **47**, R35.

FASEL, U., KUTZ, J.N., BRUNTON, B.W. & BRUNTON, S.L. 2022 Ensemble-SINDy: robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proc. R. Soc. A* **478**, 20210904. https://doi.org/10.1098/rspa.2021.0904

FISHER, D.M. & ROGERS, B.N. 2017 Two-fluid biasing simulations of the large plasma device. *Phys. Plasmas* **24**, 022303.

FREZAT, H., LE SOMMER, J., FABLET, R., BALARAC, G. & LGUENSAT, R. 2021 A posteriori learning of quasi-geostrophic turbulence parametrization: an experiment on integration steps. arXiv:2111.06841.

GARBET, X., IDOMURA, Y., VILLARD, L. & WATANABE,T. H. 2010 Gyrokinetic simulations of turbulent transport. *Nucl. Fusion* **50**, 043002.

GIANNAKIS, D. & MAJDA, A.J. 2012 Nonlinear laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl Acad. Sci. USA* **109** (7), 2222–2227.

HARTWIG, Z.S., VIEIRA, R.F., SORBOM, B.N., BADCOCK, R.A., BAJKO, M., BECK, W.K., CASTALDO B., CRAIGHILL, C.L., DAVIES, M., Estrada, J., *et al.* 2020 VIPER: an industrially scalable high-current high-temperature superconductor cable. *Supercond. Sci. Technol*. **33**, 11LT01.

HASEGAWA, A. & MIMA, K. 1978 Pseudo-three-dimensional turbulence in magnetized non-uniform plasma. *Phys. Fluids* **21** (1), 87–92.

HASEGAWA, A. & WAKATANI, M. 1987 Self-organization of electrostatic turbulence in a cylindrical plasma. *Phys. Rev. Lett*. **59** (14), 1581.

KAPTANOGLU, A.A., MORGAN, K.D., HANSEN, C.J. & BRUNTON, S.L. 2021 Physics-constrained, low-dimensional models for magnetohydrodynamics: first-principles and data-driven approaches. *Phys. Rev.* E **104**, 015206.

KUTZ, J.N., BRUNTON, S.L., BRUNTON, B.W. & PROCTOR, J.L. 2016 *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. SIAM.

MESSENGER, D.A. & BORTZ, D.M. 2021 Weak SINDy for partial differential equations. *J. Comput. Phys*. **443**, 110525.

MEZIC, I. 2005 Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn*. **41** (1–3), 309–325.

MEZIC, I. 2013 Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech*. **45**, 357–378.

MOSELEY, B., MARKHAM, A. & NISSEN-MEYER, T. 2020 Solving the wave equation with physicsinformed deep learning. arXiv:2006.11894.

MOSELEY, B., NISSEN-MEYER, T. & MARKHAM, A. 2020 Deep learning for fast simulation of seismic waves in complex media. *Solid Earth* **11**, 1527–1549.

RAISSI, M., PERDIKARIS, P. & KARNIADAKIS, G.E. 2019 Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys*. **378**, 686–707.

REINBOLD, P.A.K., GUREVICH, D.R. & GRIGORIEV, R.O. 2020 Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Phys. Rev.* E **101**, 010203.

ROWLEY, C.W., MEZIC, I., BAGHERI, S., SCHLATTER, P. & HENNINGSON, D.S. 2009 Spectral analysis of non-linear flows. *J. Fluid Mech*. **645**, 115–127.

RUDY, S.H., BRUNTON, S.L., PROCTOR, J.L. & KUTZ, J.N. 2017 Data-driven discovery of partial differential equations. *Sci. Adv*. **3** (4), e1602614.

SCHAEFFER, H. & MCCALLA, S.G. 2017 Sparse model selection via integral terms. *Phys. Rev.* E **96**, 023302.

SCHMID, P.J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech*. **656**, 5–28.

SCHMIDT, M.D., VALLABHAJOSYULA, R.R., JENKINS, J.W., HOOD, J.E., SONI, A.S., WIKSWO, J.P. & LIPSON, H. 2011 Automated refinement and inference of analytical models for metabolic networks. *Phys. Biol*. **8** (5), 055011.

SCHMIDT, M. & LIPSON, H. 2009 Distilling free-form natural laws from experimental data. *Science* **324** (5923), 81–85.

VAN DE PLASSCHE, K.L., CITRIN, J., BOURDELLE, C., CAMENEN, Y., CASSON, F. J., DAGNELIE,V. I., FELICI, F., HO, A., VAN MULDERS, S. & JET CONTRIBUTORS 2020 Fast modeling of turbulent transport in fusion plasmas using neural networks. *Phys. Plasmas* **27**, 022310.

WAKATANI, M. & HASEGAWA, A. 1984 A collisional drift wave description of plasma edge turbulence. *Phys. Fluids* **27** (3), 611–618.

WANG, W.X., LIN, Z., TANG, W.M., LEE, W.W., ETHIER, S., LEWANDOWSKI, J.L.V., REWOLDT, G., HAHM, T.S. & MANICKAM, J., *et al.* 2006 Gyro-kinetic simulation of global turbulent transport properties in tokamak experiments. *Phys. Plasmas* **13**, 092505.

WANG, Z., BOVIK, A.C., SHEIKH, H.R. & SIMONCELLI, E.P. 2004 Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process*. **13** (4), 600–612.

ZHU, B., FRANCISQUEZ, M. & ROGERS, B.N. 2017 Global 3D two-fluid simulations of the tokamak edge region: turbulence, transport, profile evolution, and spontaneous $E \times B$ rotation. *Plasmas* **24**, 055903.