

RESEARCH ARTICLE | JANUARY 13 2026

Electron neural closure for turbulent magnetosheath simulations: Energy channels

G. Miloshevich  ; L. Vranckx ; F. N. de Oliveira Lopes ; P. Dazzi ; G. Arrò ; G. Lapenta



Phys. Plasmas 33, 012901 (2026)

<https://doi.org/10.1063/5.0300009>



Articles You May Be Interested In

Magnetospheric Multiscale measurements of turbulent electric fields in earth's magnetosheath: How do plasma conditions influence the balance of terms in generalized Ohm's law?

Phys. Plasmas (August 2023)

Observing the prevalence of thin current sheets downstream of Earth's bow shock

Phys. Plasmas (October 2021)

Intermittency at Earth's bow shock: Measures of turbulence in quasi-parallel and quasi-perpendicular shocks

Phys. Plasmas (August 2023)



Physics of Plasmas

Special Topics Open
for Submissions

[Learn More](#)



Electron neural closure for turbulent magnetosheath simulations: Energy channels

Cite as: Phys. Plasmas **33**, 012901 (2026); doi: 10.1063/5.0300009

Submitted: 30 August 2025 · Accepted: 23 December 2025 ·

Published Online: 13 January 2026



View Online



Export Citation



CrossMark

G. Miloshevich,^{1,a)}  L. Vranckx,¹  F. N. de Oliveira Lopes,¹  P. Dazzi,¹  G. Arrò,²  and G. Lapenta^{1,b)}

AFFILIATIONS

¹Centre for Mathematical Plasma Astrophysics, Department of Mathematics, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium

²Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

^{a)} Author to whom correspondence should be addressed: george.miloshevich@kuleuven.be

^{b)} Deceased, May 2024.

ABSTRACT

In this work, we introduce a non-local five-moment electron pressure tensor closure parametrized by a fully convolutional neural network (FCNN). Electron pressure plays an important role in generalized Ohm's law, competing with electron inertia. This model is used in the development of a surrogate model for a fully kinetic energy-conserving semi-implicit particle-in-cell simulation of decaying magnetosheath turbulence. We achieve this by training FCNN on a representative set of simulations with a smaller number of particles per cell and showing that our results generalize to a simulation with a large number of particles per cell. We evaluate the statistical properties of the learned equation of state, with a focus on pressure-strain interaction, which is crucial for understanding energy channels in turbulent plasmas. The resulting equation of state learned via FCNN significantly outperforms local closures, such as those learned by multi-layer perceptron (MLP) or double adiabatic expressions. We report that the overall spatial distribution of pressure-strain and its conditional averages are reconstructed well. However, some small-scale features are missed, especially for the off-diagonal components of the pressure tensor. Nevertheless, the results are substantially improved with more training data, indicating favorable scaling and potential for improvement, which will be addressed in future work.

© 2026 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0300009>

I. INTRODUCTION

Understanding and predicting the behavior of collisionless space plasmas in the near-Earth environment present a fundamental scientific challenge due to its multi-scale nature. Many of the important theoretical and numerical results can be tested in this environment. Of particular interest are the energy exchanges and dissipation in collisionless plasmas, driven by reconnection and wave-particle interactions. Turbulence drives the formation of thin current sheets that reconnect and may drive secondary reconnection processes. Thus, these phenomena are intimately linked in plasmas.¹ Figuring out how plasma is energized in such turbulent environments was highlighted as one of the critical research questions for future space missions.² These processes create extreme events posing serious risks to our infrastructure.³

The Earth's magnetosheath, which separates the bow shock from the magnetopause, is located at the interface between the Earth's magnetosphere system and the solar wind that drives space weather.

The interplay between turbulence and reconnection^{4,5} in the magnetosheath has recently been a subject of study, for instance, the influence of turbulence on the length of current sheets formed.⁶ This leaves open questions, such as understanding the interaction between magnetic reconnection and other types of waves and instabilities^{7,8} that contribute to plasma heating. Indeed, the magnetosheath is also home to a variety of modes that are driven by pressure anisotropies, such as whistler waves, which have been recently associated with driving the formation of electron magnetic holes.^{9,10} These processes were studied using ECsim¹¹ [an energy-conserving semi-implicit particle-in-cell (PIC) code], which allows moderately large domains spanning ion Larmor radius-electron inertial length scales and mass ratios $m_i/m_e \sim 100$.

PIC codes¹²⁻¹⁴ are more efficient than higher fidelity Vlasov fully kinetic simulations.¹⁵ Such simulations can only be performed for certain parameter ranges and spatial scales. A more efficient approach involves reduced order models (ROMs), such as hybrid Vlasov^{16,17} and hybrid PIC codes¹⁸ and multi-fluid codes,^{19,20} which have been

developed for global modeling of magnetospheres. By *hybrid* models, we refer to models, which have kinetic ions and fluid electrons. These models have made significant advances in our ability to represent the physics at mesoscales, typically covering MHD and sub-ion scales, but crucially omitting electron scales. Electrons are typically assumed to be polytropic or even isothermal. This is in stark contrast to the physics which takes place at microscopic scales, whereby electrons are de-magnetized in the process of collisionless magnetic reconnection,²¹ where electrons also play a key role in modifying dynamics.²² These processes contribute to the heating of electrons that is not captured in hybrid models, as confirmed by simulations of the Hermean magnetosphere carried out using fully kinetic simulations²³ when compared to much cooler electrons found in hybrid simulations.²⁴ Thus, in order to understand and predict energization of plasma, it is important to couple high-fidelity fully kinetic simulations with ROMs to properly represent the physics of both electrons and ions.

One of the potentially promising avenues involves coupling PIC codes to fluid codes in an attempt to model complex objects, such as Earth's magnetosphere, more accurately.^{25–28} This allows resolving the magnetotail with a PIC simulation in a small box coupled to a larger fluid magnetospheric simulation. This approach is not justified if the entire box is dominated by kinetic physics²⁹ and fully kinetic modeling should be considered in this case.

An alternative approach involves selecting an appropriate closure relation for electrons that is embedded within hybrid simulations. The problem of collisionless fluid closure has a long history in plasma physics, dating back to the works of Chew *et al.*,³⁰ which postulated simple double-adiabatic relationships that later became known as the CGL equations, named after the authors Chew, Goldberger, and Low. It is well known that such closures break down under moderate values of finite Larmor radius (FLR) effects due to the thermal gyration of particles around magnetic fields. A more advanced version of such anisotropic pressure closure has been proposed in a series of seminal works,^{31,32} which proposes interpolation between the trapped particle dynamics, more closely resembling the CGL-like dynamics, and passing particle dynamics, which corresponds to the Boltzmann limit where the plasma is isothermal along the field lines. The evidence for this behavior was found³³ in magnetospheric multiscale (MMS)³⁴ observations. One of the limitations of CGL-like models is the inability to deal with *agyrotropy*, unequal dispersions of the velocity distribution function (VDF) perpendicular to the local field. Enhanced *agyrotropy* is typically present near the current sheets on the scale of inertial length. Early works³⁵ have shown the importance of such features of the pressure tensor in the total electric field in guide field magnetic reconnection.

It is possible to close the fluid equations by providing an equation of state for heat flux rather than pressure, in which case pressure is modeled dynamically. This allows processes such as Landau damping to be represented, which is typically not possible within a fluid framework. Such models are usually referred to as *Landau fluids* and were introduced by Hammett and Perkins³⁶ and later developed by Passot and Sulem.³⁷ Crucially, they operate under linear response approximation. Under large to moderate guide fields, such models do indeed capture the main features of magnetic reconnection,³⁸ with some exceptions such as electron-cyclotron instability. Importantly, it is challenging to apply them in low guide fields.³⁹ Although some progress has been achieved in later works,⁴⁰ the strong system-size

dependence of the average reconnection rate observed in kinetic and hybrid simulations⁴¹ was not completely reproduced, leaving room for new developments.

More recently, machine learning (ML) approaches have been widely applied in fluid dynamics⁴² and the geophysical sciences, including plasma physics.⁴³ In Earth weather, neural subgrid closures have been successfully implemented and currently rival fully physics-based solvers.⁴⁴ A *subgrid closure* is one where small-scale processes are parametrized using approximations or a statistical model. This is a popular approach in large eddy simulations (LES) with deterministic or stochastic parametrizations developed using variational autoencoders (VAEs) and generative adversarial networks (GANs).⁴⁵ While electron pressure or heat flux closure is not equivalent to this, there are parallels that have already been explored using a relatively simple ML architecture in the case of the GEM challenge, i.e., modeling collisionless magnetic reconnection. Laperre *et al.*⁴⁶ have applied histogram gradient boosting regressor (HGBR) and multi-layer perceptron (MLP) for this task to map lower order moments such as density, velocity, and electric and magnetic field to pressure tensor and heat flux vector. They have reported difficulties reconstructing off-diagonal components related to *agyrotropy*, especially in the region bound by the reconnection X-lines.

Fluid closure in plasma has already been addressed using convolutional neural networks (CNNs), which are the backbone of image analysis in computer vision. In particular, a one-dimensional Hammett–Perkins closure was previously sought;⁴⁷ however, a single dense layer with a discrete Fourier transform (DFT) appeared to perform better. The authors argued that this was due to the availability of an analytic solution, which would not be the case in higher dimensions. CNNs were also trained to represent an analytic form of collisional Braginskii's and Hammett–Perkins closure, given a range of macroscopic input profiles.⁴⁸ More complex architectures, such as VNet, have been applied to nonlocal Vlasov–Poisson with Bhatnagar, Gross, and Krook collision operator.⁴⁹ These data-driven closures were reported to be numerically stable but costly, with potential advantages in higher dimensions. More recently, a one-dimensional Landau fluid closure was obtained using a Fourier neural operator (FNO) and reportedly outperformed the Hammett–Perkins closure.⁵⁰

In parallel, there have been approaches focusing on symbolic and sparse regression,⁵¹ where equations are extracted directly from simulation or observational data. For instance, Alves and Fiuza⁵² have extracted polytropic closures from collisionless shock simulations using a popular method referred to as SINDy (symbolic identification of nonlinear dynamics).⁵³ This method has been subsequently applied to extract heat flux closures in magnetic reconnection⁵⁴ and some electrostatic phenomena.⁵⁵ While we believe that the approach is promising and there are many unexplored avenues in symbolic regression in general,⁵⁶ this falls beyond the scope of this manuscript. In addition, there are some disadvantages to using SINDy, such as the need to choose a specific library of terms, i.e., a preselected choice of potential expressions, and limitations on expressivity due to polynomial representation.

Therefore, our goal is to learn pressure and a heat flux closure using neural networks trained on multiple simulations performed with the fully kinetic code ECsim.¹¹ Our approach differs from Laperre *et al.*⁴⁶ in that we consider a problem of magnetosheath turbulence, which is a significantly more challenging setting due to the presence of

chaotic turbulent eddies. To improve upon the state of the art, we introduce a fully convolutional neural network (FCNN) that operates on patches rather than locally and study the importance of using global closure for both pressure and heat flux. We evaluate the performance of the new closure by inspecting energy channels using a scale filtering approach and discussing both the spatial structures and overall statistics. This includes evaluating anisotropy vs β_{\parallel} plots that are well bounded by whistler and electron firehose instabilities in our simulations, as well as the learned closure.

The manuscript is organized as follows. In Sec. II, we discuss the methodology starting from the description of the dataset in Sec. II A, then in Sec. II B, we introduce the closure theory and simple adiabatic closures like CGL, also introducing pressure-strain and related quantities. In Sec. III, we discuss results consisting of neural pressure and heat flux closure in Sec. III A. Next, we characterize predicted anisotropies in Sec. III B and evaluate the energy channels in Sec. III C. We address the issue of generalization across low and high fidelity PIC simulations in Sec. III D. This is followed by a discussion in Sec. IV and a conclusion in Sec. V.

II. METHODOLOGY

Collisionless plasmas evolve according to the Vlasov equation, which can be written for each species

$$\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \frac{\partial f_s}{\partial \mathbf{x}} + \frac{e_s}{m_s} \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \cdot \frac{\partial f_s}{\partial \mathbf{v}} = 0, \quad (1)$$

where $f_s = f_s(\mathbf{x}, \mathbf{v}, t)$ stands for one-particle distribution function at position \mathbf{x} having velocity \mathbf{v} for species s , while \mathbf{E} and \mathbf{B} are self-consistent electromagnetic fields, and c is the speed of light. The Vlasov equation is solved coupled to Maxwell's equations

$$\frac{\partial \mathbf{B}}{\partial t} = -c \nabla \times \mathbf{E} \quad (2)$$

and

$$\nabla \times \mathbf{B} = \frac{1}{c} \left(4\pi \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t} \right). \quad (3)$$

Here, \mathbf{J} stands for the total current, defined as

$$\mathbf{J} = \sum_s e_s n_s \mathbf{v}_s, \quad (4)$$

where n_s corresponds to the bulk density. This deceptively simple-looking system of equations can be solved numerically using Vlasov¹⁵ or particle-in-cell codes;⁵⁷ however, due to the multiscale nature of plasma, it becomes prohibitively expensive for certain problems.

A. Simulation data

In this study, we use a 2D-3V simulation described in Arrò, Califano, and Lapenta⁵⁸ (hereafter referred to as run A) that was carried out using the energy-conserving semi-implicit particle-in-cell code ECsim,¹¹ under conditions comparable to those of Earth's magnetosheath. The main parameters of the run are tabulated in column A of Table I.

For run A, we consider a 2D square periodic box of length $L = 64 d_i$, sampled by a uniform grid with 2048^2 points. Here, $d_i = c/\omega_{pi}$ stands for the ion inertial length and $\omega_{pi} = \sqrt{4\pi e_i n_i^2/m_i}$

TABLE I. Initial simulation parameters in the run A,⁵⁸ which has 5000 particles per cell, vs runs B (of which there are 6), which have 256 particles per cell.

Simulation	A: 5000 ppccell	B: 256 ppccell
$\delta B/B$	0.71	0.59
$\delta V_i/V_A$	1.1	1.1
$\delta V_e/V_A$	1.1	1.1
β_i	5.3	5.7
β_e	1.3	1.4

stands for the ion plasma frequency. We employ 5000 particles per cell for both ions and electrons, with mass ratio $m_i/m_e = 100$ (with m_i and m_e being ion and electron masses). Particles are initialized from a Maxwellian distribution with uniform density, uniform temperature, and plasma beta (defined as $\beta_s = 8\pi n_s k_B T_s/B^2$) equal to $\beta_i = 8$ for ions and $\beta_e = 2$ for electrons. A uniform initial magnetic field $\mathbf{B}_0 = B_0 \hat{\mathbf{z}}$ is present, and turbulence is induced by random magnetic field and velocity fluctuations with wavenumbers k in the range $1 \leq k/k_0 \leq 4$ (with $k_0 = 2\pi/L$). Magnetic field and velocity fluctuations have amplitudes $\delta B_{rms}/B_0 \simeq 0.9$ and $\delta u/c_A \simeq 3.6$ (where c_A is the Alfvén speed referred to B_0). The time step used to evolve the simulation is $\Delta t = 0.05 \Omega_e^{-1}$ (with $\Omega_e = e_e B_0/m_e c$ being the electron cyclotron frequency). Time is normalized to Ω_e^{-1} and the ratio of the plasma frequency to the cyclotron frequency is $\omega_{p,i}/\Omega_i = 100$ for the ions and $\omega_{p,e}/\Omega_e = 10$ for the electrons. Additional information is provided in Arrò *et al.*,⁹ Espinoza-Troni *et al.*,¹⁰ and Arrò *et al.*⁵⁸

We have performed six supporting simulations (referred to as run B) that are initialized using very similar values of the parameters, except for $\delta B/B$, which is larger in run A from the start of the simulation and remains larger over the whole simulation time (see Table I). Each member of B starts from a new random realization of the magnetic and velocity field fluctuations defined above with the prescribed levels of $\delta B/B$. Due to computational costs, we limited the six runs to 256 particles per cell; therefore, B runs tend to be noisier. For these reasons, we can use A as a higher fidelity reference point.

Simulations A and B2 have been run until the maximum $J_{rms} = \sqrt{\langle J^2 \rangle}$ is reached, where $\langle \rangle$ indicates spatial averaging. This is typically associated with the onset of a fully developed turbulence regime. Runs B1 and B3–B6 run for longer, as can be seen from Table II.

B. Fluid closure and adiabatic invariants

Instead of solving the original system of Eqs. (1)–(3), which is computationally expensive, it is possible to switch to fluid quantities.

Defining the number density $n_s(\mathbf{x}, t) := \int d^3 v f_s(\mathbf{x}, \mathbf{v}, t)$, we may integrate Eq. (1) over $\int d^3 v$ to obtain the continuity equation

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s \mathbf{V}_s) = 0. \quad (5)$$

Integrating Eq. (1) over $\int d^3 v \mathbf{v}$ and defining $\mathbf{V}(\mathbf{x}, t) := n^{-1} \int d^3 v f(\mathbf{x}, \mathbf{v}, t) \mathbf{v}$ gives the momentum equation

$$m_s \frac{\partial}{\partial t} (n_s \mathbf{V}_s) + m_s \nabla \cdot (n_s \mathbf{V}_s \mathbf{V}_s) = n_s e_s \left(\mathbf{E} + \frac{1}{c} \mathbf{V}_s \times \mathbf{B} \right) - \nabla \cdot \mathbf{P}_s, \quad (6)$$

TABLE II. Description of the datasets that are partitioned between test/validation and training. The first column indicates the name of the run, the second column indicates the available and retained timesteps, and the next column indicates which partition these data end up in. There is also split number 2, which is used solely in Appendixes.

	Time steps [ω_{pi}]	Split 1	Split 2
PIC A	375, 400, ..., 750	Test 2	None
PIC B 1	375, 400, ..., 1200	Test	Train
PIC B 2	375, 400, ..., 700	Validation	Train
PIC B 3	375, 400, ..., 1225	Train	Test
PIC B 4	375, 400, ..., 1200	Train	Train
PIC B 5	375, 400, ..., 1050	Train	Validation
PIC B 6	375, 400, ..., 1325	Train	Train

where we have defined the pressure tensor $\mathbf{P}_s := m_s \int d^3v (\mathbf{v}_s - \mathbf{V}_s)(\mathbf{v}_s - \mathbf{V}_s) f(\mathbf{x}, \mathbf{v}_s, t)$. Neglecting the displacement current in Eq. (3), defining the bulk (average) plasma velocity as \mathbf{V} , and considering the mass ordering $m_e \ll m_i$, we get

$$\mathbf{E} - \frac{d_e^2}{n} \Delta \mathbf{E} = -\mathbf{V} \times \mathbf{B} + \frac{1}{n} (\mathbf{J} \times \mathbf{B}) + \frac{d_e^2}{n} \nabla \cdot \mathbf{P}_i - \frac{1}{n} \nabla \cdot \mathbf{P}_e + \frac{d_e^2}{n} \nabla \cdot [\mathbf{J} + \mathbf{J}\mathbf{V}] - \frac{d_e^2}{n} \nabla \cdot \left(\frac{\mathbf{J}\mathbf{J}}{n} \right), \quad (7)$$

which is known as the *generalized Ohm's law*. Here, Δ stands for Laplacian. We have written this equation in normalized units, so that spatial coordinates scale as d_i , the *ion inertial length*. The terms multiplying d_e^2 , representing the electron inertial length, are related to electron inertia effects. They play an important role in the electron diffusion region (EDR), near the X-point of reconnection events, where they compete with the influence of the electron pressure tensor \mathbf{P}_e , which is the subject of this manuscript.

If the generalized Ohm's law, Eq. (7), is coupled with the fluid or Vlasov equation [Eq. (1)] for ions, under an appropriate choice of \mathbf{P}_e , the electron pressure tensor, the system is closed. If the ion closure is also prescribed (e.g., at the level of the ion pressure tensor \mathbf{P}_i), the system is sometimes referred to as eXtended MHD (XMHD) and can be proven to be Hamiltonian,^{59,60} i.e., possesses structure preservation and even topological invariants.⁶¹

The standard collisional magnetohydrodynamic condition for \mathbf{P}_e corresponds to an adiabatic equation of state, $\mathbf{P}_e \sim n^{\gamma} \mathbf{I}$, where \mathbf{I} stands for the identity matrix. This relationship breaks down for plasmas that are not in local thermal equilibrium. Under the assumption that the magnetic field varies only weakly over a Larmor radius (an approximation that is reasonable in many smoothly varying magnetized plasmas but often violated in collisionless reconnection regions), one may treat the first and second adiabatic invariants as approximately conserved

$$\mu := \frac{mV_{\perp}^2}{2B} = \text{const.}; \quad \mathcal{J} := mV_{\parallel}L = \text{const.} \quad (8)$$

Indeed, in the absence of Hall, finite Larmor radius (FLR) effects, and heat flux, one arrives at the well-known CGL (Chew–Goldberger–Low) condition³⁰ $p_{\parallel} \sim n^3/B^2$ and $p_{\perp} \sim nB$, see, for instance, Hunana *et al.*⁶² We consider a more general closure found in seminar works of Le *et al.*,³¹ as follows:

$$\begin{aligned} \tilde{p}_{\parallel,e} &= \frac{2\xi\alpha}{2\xi\alpha + 1} \frac{\pi\tilde{n}^3}{6\tilde{B}^2} + \frac{2}{2 + \xi\alpha} \tilde{n}, \\ \tilde{p}_{\perp,e} &= \frac{\xi\alpha}{\xi\alpha + 1} \tilde{n}\tilde{B} + \frac{1}{1 + \xi\alpha} \tilde{n}, \end{aligned} \quad (9)$$

where $\alpha = n^3/B^2$, and for any quantity Q , $Q_* = Q/Q_{\infty}$, where Q_{∞} is the value of Q , the reference region away from current sheets. In this model, the electron equation of state smoothly interpolates between a regime where trapped electrons, with strongly reduced parallel heat conduction, approximately follow CGL-like double-adiabatic scalings, and a regime where the passing electrons, which retain efficient parallel heat transport along the field, exhibit an almost isothermal (Boltzmann-like) response.³³ The transition between the regions is controlled by the parameter α . It was originally applied in simpler current sheet conditions. We chose to adapt the model (9) by fitting a multiplier ξ in front of the parameter α to better match the data.

It is possible to obtain more accurate but more expensive 10-moment closures, where the constituent relation is for the heat flux tensor, and thus, pressure is evolved dynamically. Here, we present a version of the heat flux vector under the assumption of a gyrotropic pressure tensor⁶³

$$\begin{aligned} \frac{dp_{\parallel,e}}{dt} &= -p_{\parallel,e} \nabla \cdot \mathbf{u}_e - 2p_{\parallel,e} \mathbf{b} \cdot \nabla \mathbf{u}_e \cdot \mathbf{b} \\ &\quad - \nabla \cdot (q_{\parallel,e} \mathbf{b}) + 2q_{\perp,e} \nabla \cdot \mathbf{b}, \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{dp_{\perp,e}}{dt} &= -2p_{\perp,e} \nabla \cdot \mathbf{u}_e + p_{\perp,e} \mathbf{b} \cdot \nabla \mathbf{v}_e \cdot \mathbf{b} \\ &\quad - \nabla \cdot (q_{\perp,e} \mathbf{b}) - q_{\perp,e} \nabla \cdot \mathbf{b}. \end{aligned}$$

In general, at EDR, the electron pressure tensor tends to develop agyrotropy. Following Swisdak,⁶⁴ we define *agyrotropy* as

$$A = \frac{p_{xy}^2 + p_{xz}^2 + p_{yz}^2}{p_{\perp}^2 + 2p_{\parallel}p_{\perp}}. \quad (11)$$

Agyrotropy tends to be present in conjunction with finite Larmor radius (FLR) effects.

C. Neural closure

The main goal in this manuscript is to seek a mapping for the electron pressure tensor

$$\mathbf{P} = \mathbf{P}_{\theta}(n, \mathbf{V}_e, \mathbf{E}, \mathbf{B}) \quad (12)$$

in terms of lower order moments and neural network hyperparameters θ . We make no assumption concerning the orientation of the magnetic field to achieve closure, which also works in regions of magnetic field reversals. To achieve such closure, we employ two alternative approaches.

The multi-layer perceptron (MLP)⁶⁵ is an approach identical to that of Laperre *et al.*,⁴⁶ where the input fields are fed pointwise into the layers of the fully connected neural network (see Fig. 1)

$$\hat{y}_{\theta} = f^{(L)} \left(W^{(L)} f^{(L-1)} \left(\dots f^{(1)} \left(W^{(1)} \mathbf{x} + \mathbf{b}^{(1)} \right) \dots \right) + \mathbf{b}^{(L)} \right), \quad (13)$$

where \mathbf{x} are the inputs to the network, $W^{(l)}$ are referred to as weights and \mathbf{b} as biases, while f stands for nonlinear activation functions and \hat{y}

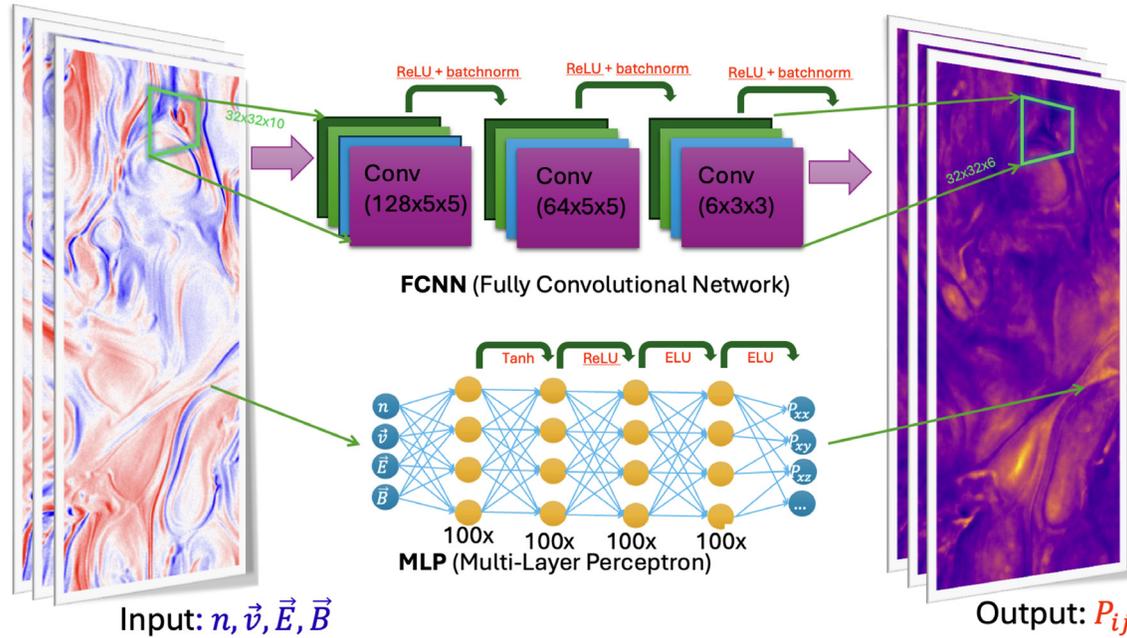


FIG. 1. Input/output and architecture of the fully convolutional neural network (FCNN), top, and multi-layer perceptron (MLP), bottom. The images on the left are being fed into both architectures using either a patch-based approach, as indicated by a green frame in the case of FCNN, or a point-based approach, as indicated by a green arrow in the case of MLP. On the right, the output pressure tensor is plotted. The internal architecture of FCNN consists of three convolutional layers with (number of channels, kernel dimension 1, kernel dimension 2) indicated on top. The green arrows indicate the application of activation functions and batch normalization. MLP consists of four layers, each with 100 neurons.

is the prediction made by the network. The operation in Eq. (13) is usually referred to as the *forward pass*.

The fully convolutional neural network (FCNN) is an architecture consisting solely of convolutional layers (see Fig. 1). Each convolutional layer performs, for each pixel of the image, a multiplication of its value and neighboring values with the corresponding elements of the kernel matrix. Each layer is equipped with a certain number of such kernels that are concatenated together, thus producing a tensor W whose outer dimensions are referred to as the channel or filter dimensions, whereas the inner indices with respect to which the inner product is taken correspond to spatial dimensions. We provide a mathematical expression for a single convolutional layer with C_{in} channels as inputs, and kernel size $(k_w \times k_h)$, as follows:

$$Y_{o,i,j} = \sum_{c=1}^{C_{in}} \sum_{u=-k_h/2}^{k_h/2-1} \sum_{v=-k_w}^{k_w-1} W_{o,c,u,v} \times X_{c,i+u,j+v} + b_o. \quad (14)$$

The choice of FCNN ensures full translation invariance because only the neighbors of each node are connected. This also makes FCNN much more efficient and lightweight than connecting every point to every other point, which would be impractical and lead to overfitting. To summarize, due to its geometry, FCNN allows us to feed it patches or entire images rather than points, as in the case of MLP. The way each layer is padded ensures that its output dimension matches its input dimension. In principle, the forward pass can still be formally represented by the operation in Eq. (13), but not all weights are allowed.

The weights and biases are obtained by training the networks on the data, minimizing the mean squared error (MSE) between the predicted values and the actual data (ground truth)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_{i,\theta})^2. \quad (15)$$

The choice of the loss function is motivated by the fact that MSE corresponds to cross-entropy over a continuous Gaussian variable. We have verified that moments, such as pressure, indeed follow distributions that are close to Gaussian in the data. In principle, more advanced loss functions can be considered in the future to refine the optimization objective. To minimize MSE in Eq. (15), the procedure of *backpropagation* is applied, which is essentially an optimization problem that gradually adjusts the weights of the network according to the gradients of MSE (gradient descent). To evaluate the quality of predictions, R^2 determination score is computed on validation and test sets

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_{i,\theta})^2}{\sum_i (y_i - \bar{y})^2}, \quad (16)$$

i.e., R^2 measures how large MSE is compared to the typical variance in the data.⁵⁶ Thus, $R^2 = 1$ corresponds to perfect prediction, while $R^2 < 0$ indicates that the error exceeds the typical variance in the data. We note that here R^2 is an ensemble metric that evaluates the quality of predictions pointwise. As such, it is not affected by the image's spatial integrity and is strongly affected by noise in the PIC data, which

affect small scales. We have experimented with other scores such as the Pearson correlation coefficient and the structural similarity index measure (SSIM), but we found that they often followed the same trends as R^2 ; thus, we decided to keep the study simpler by focusing on one score and visual inspection.

D. Datasplit

When applying machine learning (ML) to physical science, it is crucial to properly split the data into training, validation, and testing sets to prevent data contamination. The training set is the set on which backpropagation and weight and bias optimization are performed. The validation set is the set used to optimize hyperparameters, such as the number of layers and when to stop training. The test set is a separate dataset used to evaluate the performance of several successful ML architectures. There are four levels of data split that can be defined, ranked according to the degree of difficulty. *Random* split involves randomly splitting the data into the aforementioned sets. This approach would be highly problematic for evaluation, as spatially correlated data may be mixed in both training and testing, and the resulting metrics would not be relevant for generalizing to a future state or a new run. *Chronological* split involves taking chunks of temporal data and distributing them accordingly across training, testing, and validation. This is a better approach, but it is most suitable when a stationary regime has been achieved. Nevertheless, there is still no guarantee that satisfactory performance on a chronological split implies that the network will generalize on a new run. *Initialization* split uses an ensemble of runs that share the same governing parameters, and differences arise only from random initialization, turbulence seeds, or noise. This is the approach we take in the current study, where the runs are labeled B1, B2, B3, B4, B5, and B6 (see Table II and the corresponding timeshots used). This way, we can confirm that the methodology generalizes across new initial conditions.

Finally, the most challenging from an ML perspective is the *out-of-distribution* (OOD) split. This implies changing the characteristic parameters, such as β_e and $\delta B/B$, across validation training/testing (distributing runs with different values of these parameters in training and test sets). From Table I, we see that run A has larger values of $\delta B/B$. The difference appears relatively mild, although further analysis reveals that this difference persists over time.

E. Pressure-strain and scale filtering

Here, we review the quantities of interest (QOIs) useful for estimating the transfer of energy between flow, thermal, and electromagnetic fields that have been extensively studied in the past by Yang *et al.*⁶⁷ and Matthaeus *et al.*⁶⁸ Using the equations of motion (1)–(3), one can arrive at the following set:

$$\partial_t E_s^f + \nabla \cdot (E_s^f \mathbf{V}_s + \mathbf{P}_s \cdot \mathbf{V}_s) = (\mathbf{P}_s \cdot \nabla) \cdot \mathbf{V}_s + \mathbf{J}_s \cdot \mathbf{E}, \quad (17a)$$

$$\partial_t E_s^{th} + \nabla \cdot (E_s^{th} \mathbf{V}_s + \mathbf{q}_s) = -(\mathbf{P}_s \cdot \nabla) \cdot \mathbf{V}_s, \quad (17b)$$

$$\partial_t E^m + \frac{c}{4\pi} \nabla \cdot (\mathbf{E} \times \mathbf{B}) = -\mathbf{J} \cdot \mathbf{E}, \quad (17c)$$

where we follow the standard notation for fluid kinetic energy $E_s^f := n_s V_s^2/2$, thermal energy $E_s^{th} := m_s \int d^3v (\mathbf{v} - \mathbf{V}_s)^2 f_s(\mathbf{x}, \mathbf{v}, t)/2$, electromagnetic energy $E^m := (B^2 + E^2)/(8\pi)$, current $\mathbf{J}_s := n_s e_s \mathbf{V}_s$,

and heat flux vector $\mathbf{q}_s := m_s \int d^3v (\mathbf{v} - \mathbf{V}_s)^2 (\mathbf{v} - \mathbf{V}_s) f_s(\mathbf{x}, \mathbf{v}, t)/2$. The terms containing spatial divergences on the l.h.s. are responsible for spatial redistribution, whereas the terms of the r.h.s. can be considered as sources and are a major focus. It is convenient to represent the pressure-strain term

$$\begin{aligned} -(\mathbf{P}_s \cdot \nabla) \cdot \mathbf{V}_s &= -p \delta_{ij} \partial_j V_i^s - (P_{ij}^s - p_s \delta_{ij}) \partial_j V_i^s \\ &= -p_s \theta_s - \Pi_{ij}^s D_{ij}^s =: -p_s \theta_s - PiD_s, \end{aligned} \quad (18)$$

where $p_s := P_{ii}^s/3$ is the isotropic part, $\Pi_{ij}^s := P_{ij}^s - p_s \delta_{ij}$ is the deviatoric part, $\theta_s := \nabla \cdot \mathbf{V}_s$ is the compressible pressure dilation, and $D_{ij}^s := (\partial_i V_j^s + \partial_j V_i^s)/2 - \theta_s \delta_{ij}/3$. The last term of Eq. (18) is referred to as *pressure-strain* interaction or “Pi-D.”

The connection between coherent structures and energy conversion⁶⁷ can be examined by comparing the spatial distribution of “Pi-D” with that of the symmetric velocity stress, the vorticity, and the current density. As scalar measures, we use the normalized quadratic invariants. For the symmetric, traceless rate-of-strain tensor D_{ij} , we define

$$Q_D^s = \frac{1}{2} D_{ij}^s D_{ij}^s / \langle 2D_{ij}^s D_{ij}^s \rangle, \quad (19)$$

for the vorticity $\boldsymbol{\omega} = \nabla \times \mathbf{v}$

$$Q_\omega^s = \frac{1}{4} \boldsymbol{\omega}_s^2 / \langle \boldsymbol{\omega}_s^2 \rangle, \quad (20)$$

and for the current density

$$Q_J^s = \frac{1}{4} \mathbf{J}_s^2 / \langle \mathbf{J}_s^2 \rangle. \quad (21)$$

The purpose of this is to perform conditional averages of *PiD* thresholding above a certain value, which corresponds to identifying coherent structures in the simulation. For scale filtering performed on the quantities, see Appendix A.

As mentioned in Sec. II A runs B1–B6 have only 256 particles per cell. This generally results in a noisy electric field. To reduce the effects of particle noise, we have applied a boxcar filter to smooth all fields across all runs and times using a 4×4 kernel. This procedure is followed by downsampling from the full-resolution 2048×2048 to 512×512 .

III. RESULTS

We obtain pressure and heat flux closure by training the neural networks on data split 1, as shown in Table II. We evaluate the performance by comparing the predicted pressure and derived quantities, such as pressure-strain and agyrotropy, to the *ground truth*, i.e., the actual values in run B1. In addition, we provide overall statistics of the comparisons, such as the determination score R^2 . In what follows, we focus on electron closure; thus, when the species index is not specified, we imply $s = e$ for electrons.

A. Pressure and heat flux

To provide a baseline model for anisotropic closure, we take model (9) developed by Le *et al.*³¹ for parallel p_{\parallel} and perpendicular p_{\perp} pressure, while introducing a fitting parameter ζ to better match the simulations. Since in our domain we have multiple current sheets driven by turbulence rather than a single one,³¹ we choose to

TABLE III. R^2 values of different closure models for parallel, perpendicular, and pressure tensor components. Linear fits of symbolic expressions discussed in Sec. II B with fits performed on the CGL [related to Eq. (8)], generalized Le-Egedal (9). The fits are performed, showing R^2 for parallel and perpendicular pressures.

Model	CGL	Symbolic	MLP	FCNN	B2-trained
p_{\parallel}	-0.75	0.53	0.58	0.75	0.73
p_{\perp}	0.64	0.67	0.61	0.82	0.76
P_{xx}	0.59	0.82	0.74
P_{yy}	0.57	0.82	0.74
P_{zz}	0.57	0.80	0.72
P_{xy}	-0.04	0.37	0.18
P_{xz}	0.01	0.41	0.17
P_{yz}	0.00	0.39	0.19
q_x	0.03	0.20	...
q_y	0.04	0.23	...
q_z	0.06	0.30	...

normalize asymptotic \tilde{B} and \tilde{n} in Eq. (9) to spatially averaged values. We then fit parallel and perpendicular pressure with a corresponding single value of ζ , which is thus a global quantity.

This model is referred to as “symbolic” and is tabulated in Table III, see discussion in Sec. II B. It can be seen that it achieves the

same R^2 for p_{\perp} as the simpler double-adiabatic CGL model; however, CGL yields very poor p_{\parallel} performance with a negative R^2 . A negative determination score implies that the model is making predictions that are farther from the ground truth than the typical data variance, indicating that it is not only incorrect conditional on the inputs (n, B) but also unconditionally. Note that here we are referring to a symbolic closure³¹ rather than a machine learning based one, which takes more inputs.

In Table III, we also compare the results of the symbolic fit to those of the multi-layer perceptron (MLP) and fully convolutional neural network (FCNN) introduced in Sec. II C with the pipeline graphically represented in Fig. 1. From Table III, we see that evaluation of MLP and symbolic model (9) results in comparable p_{\parallel} and p_{\perp} . In contrast, FCNN outperforms MLP on all metrics, yielding relatively good $R^2 \gtrsim 0.8$ for diagonal components of the pressure tensor. FCNN yields below average $R^2 \sim 0.4$ determination score for off-diagonal components, which significantly outperforms MLP $R^2 \sim 0$.

To provide spatial characteristics of the neural closures, in Fig. 2, we plot pressure components P_{xx} and P_{xy} from a subset of the simulation at $t = 500\omega_{pi}^{-1}$, focusing on an island chain that has just undergone reconnection. Fully developed turbulence is typically identified by the maximum of the root mean square (RMS) current, which occurs at $t = 550\omega_{pi}^{-1}$, indicating we are near this regime. We see that FCNN prediction for P_{xx} in Fig. 2(b) matches Fig. 2(a) better than MLP prediction in Fig. 2(c). For instance, a ridge just outside of the main magnetic island at $(15d_i, 37d_i)$ is missed by the MLP.

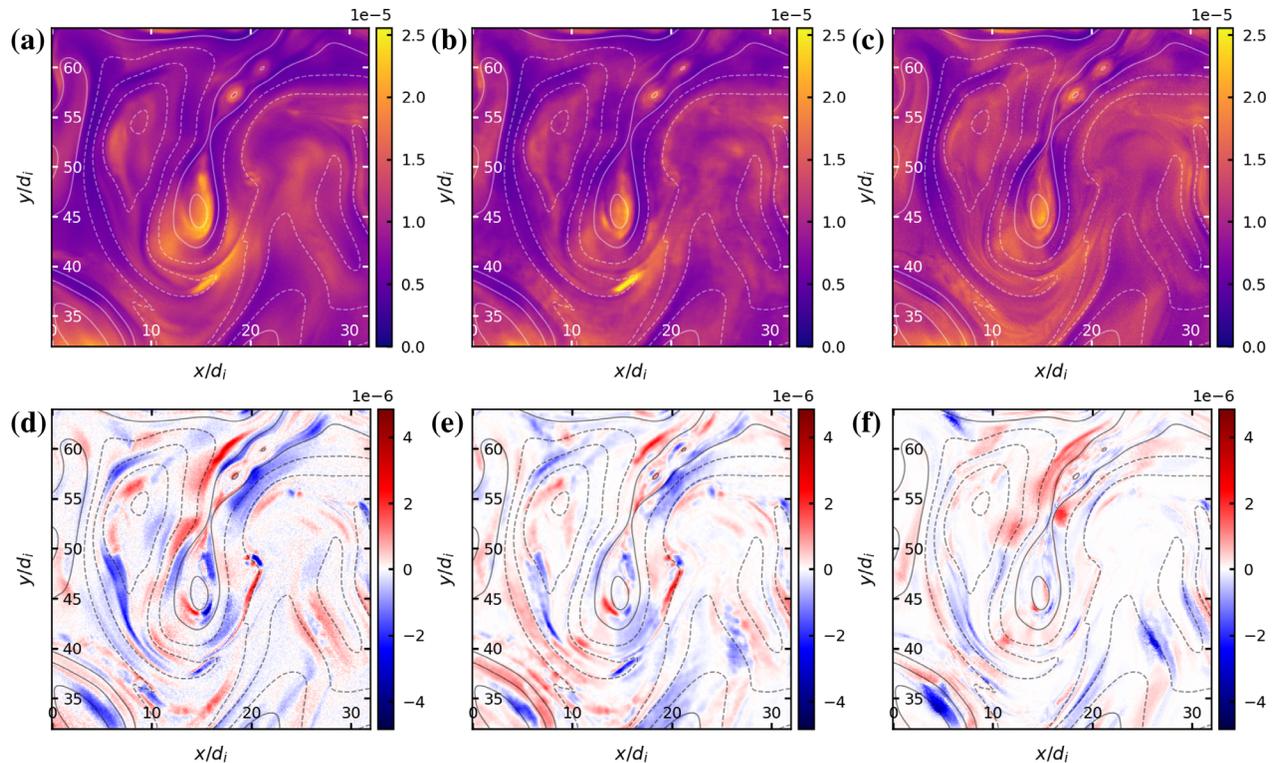


FIG. 2. Evaluation of the MLP and FCNN closures on a frame $t = 500\omega_{pi}^{-1}$, with run B1 serving as a test set (see Table II). Rows correspond to different pressure tensor components (P_{xx} , and P_{xy}), while columns correspond to (a) and (d) ground truth, (b) and (e) FCNN predictions, and (c) and (f) MLP predictions. Each quantity corresponds to the pressure tensor components, with the corresponding color bar on the right. To provide a reference, we add contours of A_z , which is equivalent to the flux function in 2D.

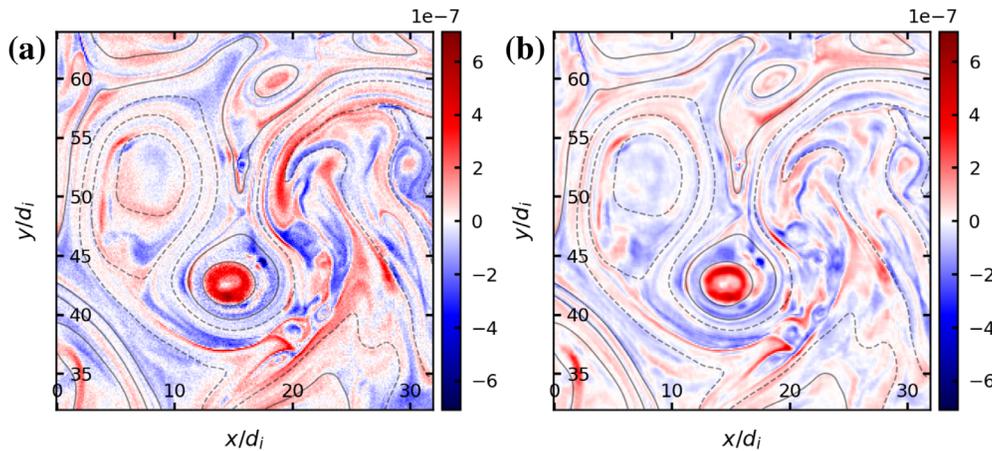


FIG. 3. Plot of z-component of heat flux q_z at $t = 825 \omega_{pi}^{-1}$ for (a) ground truth and (b) FCNN prediction. Heat flux vector plots are equipped with a corresponding color bar on the right. To provide a reference, we add contours of A_z , which is equivalent to the flux function in 2D.

In addition, FCNN is more faithful to the overall structures, but we note small-scale vapor-like artifacts that are not present in the original P_{xx} or the MLP prediction. In contrast, MLP prediction appears smoother. The scale of the artifacts is below d_i , and it does not consistently reach the scale of the grid.

FCNN prediction in Fig. 2(e) for P_{xy} captures overall features of the ground truth in Fig. 2(d) significantly better than MLP in Fig. 2(f). For instance, the positive P_{xy} at the separatrix at $(16d_i, 53d_i)$ is more accurately reproduced with FCNN, while MLP tends to predict sign reversal incorrectly. FCNN is more accurate at reproducing the signs of the ridges around the magnetic island $(15d_i, 37d_i)$. Nevertheless, one cannot help but notice certain small-scale irregularities in FCNN prediction, which can also be described as vapor-like noise. These features are consistent with overall scores reported in Table III.

We complete this subsection by reporting on the spatial structures of FCNN predicted heat flux q_z in Fig. 3(b). Despite the relatively low R^2 score reported in Table III, we see that FCNN captures main structures, for instance, counter streams of q_z inside magnetic islands, and even structures around the ridges. Upon closer inspection, we see the same vapor-like noise, which likely contributes to a low R^2 score. In Table III, we have also included comparison of overall R^2 obtained with FCNN and with MLP for \mathbf{q} . We see that MLP is much inferior at reconstructing the heat flux closure.

B. Characteristics of anisotropies

We utilize the synthetic electron pressure predicted by FCNN and MLP and apply it along with the predictor variables such as lower order moments n_e , \mathbf{V}_e , \mathbf{E} , \mathbf{B} to compute derived quantities. We plot ground truth agyrotropy defined in Eq. (11) in Fig. 4(a) at $t = 500\omega_{pi}^{-1}$ at the same location as Fig. 2. We observe that agyrotropy is strongest at the X point and along the separatrices. It also takes large values near the ridge surrounding the principal magnetic island in the island chain. This general structure is replicated in the FCNN closure estimated agyrotropy in Fig. 4(b). However, MLP closure hardly predicts any agyrotropy in Fig. 4(c), which is consistent with $R^2 \sim 0$ for off-diagonal elements as indicated in Table III.

As is well known, anisotropies lead to microinstabilities, such as the whistler instability,⁷¹ which constrain anisotropies. This is usually illustrated by plotting⁷² the anisotropy T_{\parallel}/T_{\perp} vs β_{\parallel} as we do in Fig. 5. Figure 5(a) corresponding to ground truth shows that the anisotropy parameter space is constrained from above by whistler instability and from below by electron firehose instability, with some traces outside those ranges. Concurrently, FCNN in Fig. 5(b) predicts a similar distribution respecting the instability thresholds, but it appears to reduce the variance of T_{\parallel}/T_{\perp} anisotropy somewhat. The pressure computed by MLP has an even smaller range, as shown in Fig. 5(c), which bolsters previous results comparing the two architectures. The fact that we observe lower values of anisotropies for FCNN and especially MLP is most likely due to the fact that there is regression to the mean, i.e., the neural network has a tendency to predict fewer extreme values since it was trained to optimize mean square error.

C. Energy channels

Energy channels discussed in Sec. II E express the conversion of flow (17a) into thermal energy (17b) signified by the pressure-strain term as well as more conventional Ohmic dissipation that converts electromagnetic energy (17c) into flow (17a). First, we are going to investigate scale-to-scale k-filtering defined in Appendix A and applied to pressure-strain $PS(k > k_c)$ (18), i.e., with a high-pass filter at scale k_c . We plot $PS(k > k_c)$ in Fig. 6(a), which demonstrates that overall, both FCNN and MLP capture the general distribution of pressure-strain over the scales, with FCNN shadowing accurately the quantity and MLP underestimating it by a factor of 2.

Since pressure-strain quantifies the particle heating, it is of interest to investigate whether it peaks on coherent structures, as was done in Yang *et al.*⁶⁷ For this aim, we compute the average incompressible portion of pressure-strain $\langle PiD|Q > Q^* \rangle$, where Q stands for any of the three quantities in Eqs. (19)–(21), each represented by dashed, dotted-dashed, and solid lines in Fig. 6(b), which compares ground truth to FCNN and Fig. 6(c) for MLP. We note the similarity of $\langle PiD|Q > Q^* \rangle$, which tends to be larger for larger thresholds, consistent with Yang *et al.*⁶⁷ However, in this case, the same trend is also observed for Q_j conditionals, indicating the association of current

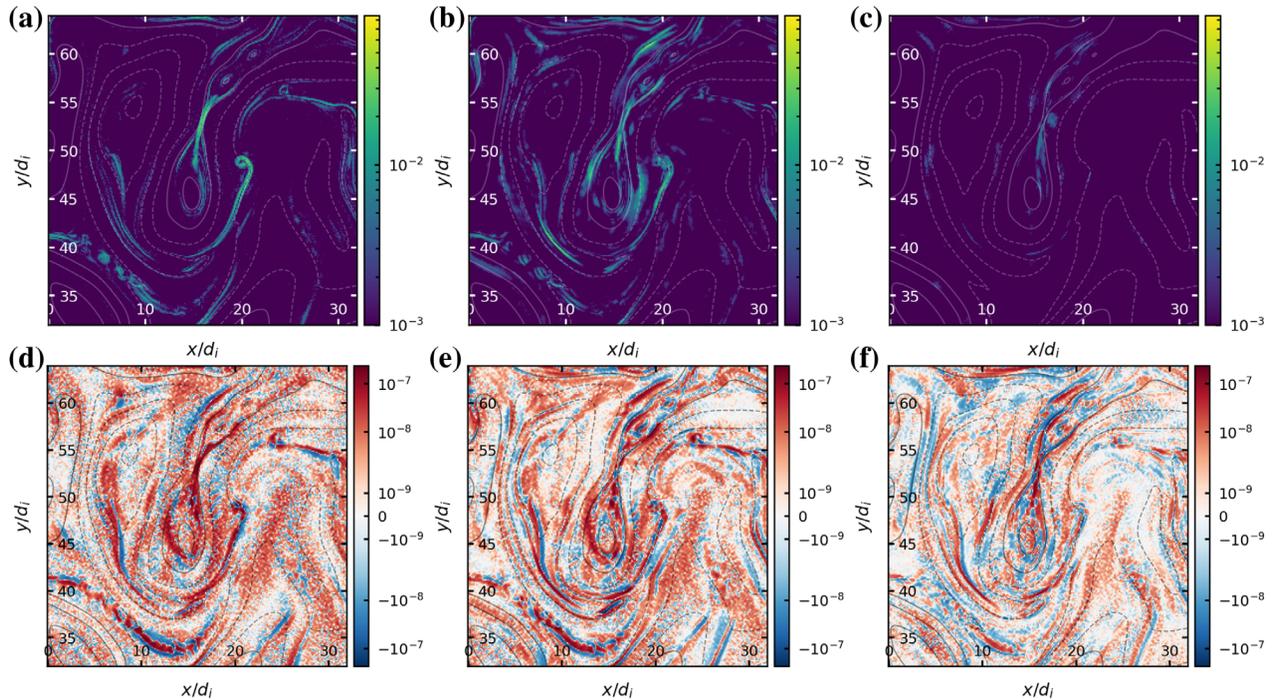


FIG. 4. Plot of agyrotropy [Eq. (11)] and incompressible pressure-strain PiD at $t = 500 \omega_{di}^{-1}$ for (a) agyrotropy ground truth, (b) FCNN prediction of agyrotropy, (c) MLP prediction of agyrotropy, (d) PiD ground truth, (e) FCNN prediction of PiD , and (f) MLP prediction of PiD . To provide a reference, we add contours of A_z , which is equivalent to the flux function in 2D.

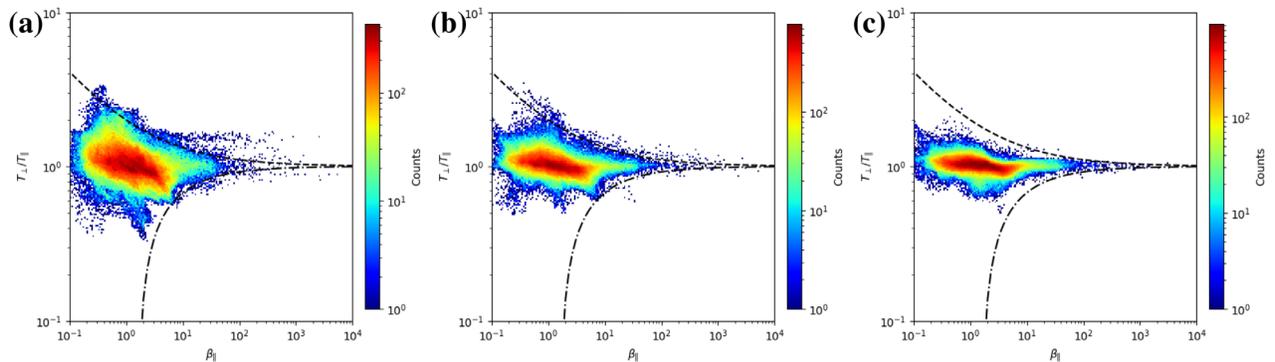


FIG. 5. Temperature anisotropy vs β_{\parallel} plots (a.k.a. Brazil plot) histograms with counts represented on the rainbow colormap for (a) ground truth. (b) FCNN prediction. (c) MLP prediction. The dashed line corresponds to the onset of whistler instability, while the dotted-dashed line corresponds to the onset of the firehose instability. The functional form of the curves is given by the expression $T_{\perp}/T_{\parallel} = 1 + A/\beta_{e\parallel}^B$, where we have taken $A = 1, B = 0.49$ for whistlers⁶⁹ (dashed line) and $A = 1.32, B = 0.61$ for electron firehose⁷⁰ (dotted-dashed line). These parameters correspond to specific growth rates, $\gamma = 0.1 \Omega_e$.

sheets and heating in our simulations. The comparison reveals qualitatively similar behavior for FCNN closure but rather poor results for MLP closure, which completely underestimates conditionals of $\langle PiD | Q > Q^* \rangle$.

D. Ablation study and generalization

Our aim here is twofold. First, we would like to perform an *ablation study*, i.e., remove certain features (inputs such as $n_e, \mathbf{V}_e, \mathbf{E}, \mathbf{B}$ from $\mathbf{P}_e = \mathbf{P}_{NN}(n_e, \mathbf{V}_e, \mathbf{E}, \mathbf{B})$ and train from scratch the FCNN.

This gives us a collection of models that we refer to as *default* := $\mathbf{P}_{NN}(n_e, \mathbf{V}_e, \mathbf{E}, \mathbf{B})$, *noE* := $\mathbf{P}_{NN}(n_e, \mathbf{V}_e, \mathbf{B})$, *Jtot* := $\mathbf{P}_{NN}(n_e, \mathbf{J}, \mathbf{E}, \mathbf{B})$, and *JtotnoE* := $\mathbf{P}_{NN}(n_e, \mathbf{J}, \mathbf{B})$ and allows us to study the influence of each feature in predicting our target \mathbf{P} . Second, we would like to study the generalization of the neural network trained, validated, and tested on B1-B6 to run A (see Table II). To this end, we plot the results of the ablation/generalization study in Fig. 7. It is organized into three panels: Fig. 7(a) applies the test set B1, Fig. 7(b) applies the test set A, and Fig. 7(c) illustrates the difference between the two.

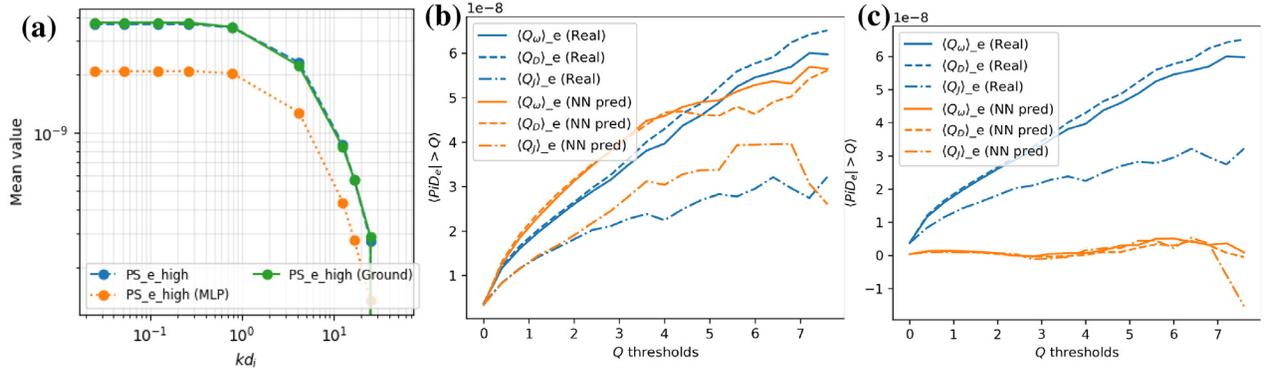


FIG. 6. (a) Total pressure-strain PS with high pass filter k^* wavenumber applied. On the x-axis, we plot k^* . The blue dashed curve shows the FCNN prediction, the green solid curve shows the ground truth, and the orange dotted curve shows the MLP prediction. (b) and (c) Average incompressible pressure-strain PiD_e , conditional to values of quantities Q_{ω} , Q_D , and Q_j above a certain threshold [see Eqs. (18) and (19)]—dashed line, (20)—solid line, and (21)—dotted-dashed line. This quantity is computed for the ground truth (blue curves) and neural network prediction (orange curves). Note that in (b), the orange curve corresponds to the predictions made with the help of FCNN, while in (c), it corresponds to the predictions made with the help of MLP.

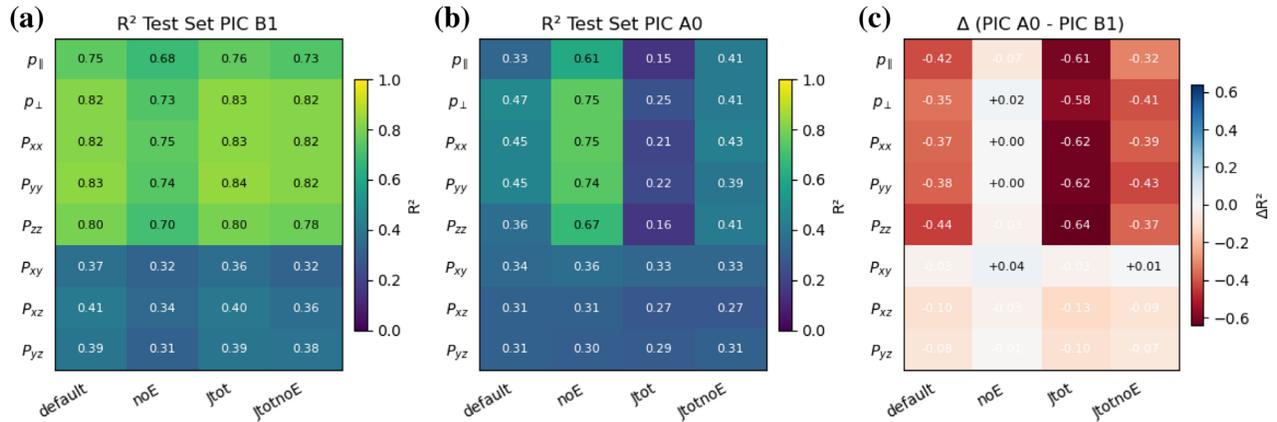


FIG. 7. Ablation study of input configurations for the FCNN. Panels show R^2 score tables for (a) test set PIC B1, (b) test set PIC A, and (c) their difference $\Delta = A - B1$ for each predicted quantity (rows) and model variant (columns). Each panel consists of a grid with rows labeled by specific quantity that is predicted, such as $p_{\parallel}, p_{\perp}, \dots$ while the columns are labeled by different models, which are distinguished by their inputs: *default* = (n_e, V_e, E, B) , *noE* = (n_e, V_e, B) , *Jtot* = (n_e, J, E, B) , and *JtotnoE* = (n_e, J, B) . Warmer colors indicate higher R^2 , see the color bar; in the difference panel on the right, red (blue) denotes improvement (degradation) on PIC B1 relative to PIC A.

The conclusion from Fig. 7(a) is that test set B1 is less sensitive toward the collection of ablation models. The worst-performing (marginally) model is *noE*, for instance, the *default* results in $R^2 = 0.82$ for p_{\perp} , while *noE* yields 0.73. Likewise, the *default* model yields $R^2 = 0.75$ for p_{\parallel} , while *noE* gives 0.68. The effect is strongest for diagonal components. This indicates that the electric field contains useful information for predicting the heating of plasma, albeit marginally. On the other hand, when the same set of models is applied to test set A, a different picture is observed in Fig. 7(b). The most noticeable is the complete reversal of the performance of the *noE* model, which turns out to be the most performant model for the diagonal part of the pressure tensor. For instance, it yields $R^2 = 0.75$ for p_{\perp} , while *default* only results in $R^2 = 0.47$. Similarly, *default* yields $R^2 = 0.61$ for p_{\parallel} , while *noE* yields $R^2 = 0.61$. The worst-performing model is *Jtot* with $R^2 = 0.15$ for p_{\parallel} and $R^2 = 0.25$ for p_{\perp} . The off-diagonal components of the pressure tensor are largely unaffected. The conclusion that can also be

drawn from Fig. 7(c) is that *noE* is the most statistically robust model; therefore, we use it for the remainder of the manuscript. We provide the following explanation for why *noE* model performs better. Run A has a much less noisy electric field than runs B1–B6. This means there is a significant shift in the distribution of E when comparing A and B. The currently used 4×4 box filter was insufficient to address this problem; thus, we propose dropping E from the predictors.

Next, we investigate the fidelity of the spatial structures of the FCNN closure of *noE* model in Fig. 8. The ground truth that comes from run A consists of several current sheets that have already become unstable. The choice of the time snapshot also occurs near the maximum J_{rms} , like in Fig. 2. The most interesting one is located at $(6d_i, 35d_i)$ and leads to enhancement of heating as can be inferred from large values of P_{xx} in Fig. 8(a). In Fig. 8(b), strong positive and negative values of P_{xy} are seen at the right separatrix, and a negative enhancement of P_{xy} is seen just north of the X point. The inspection of

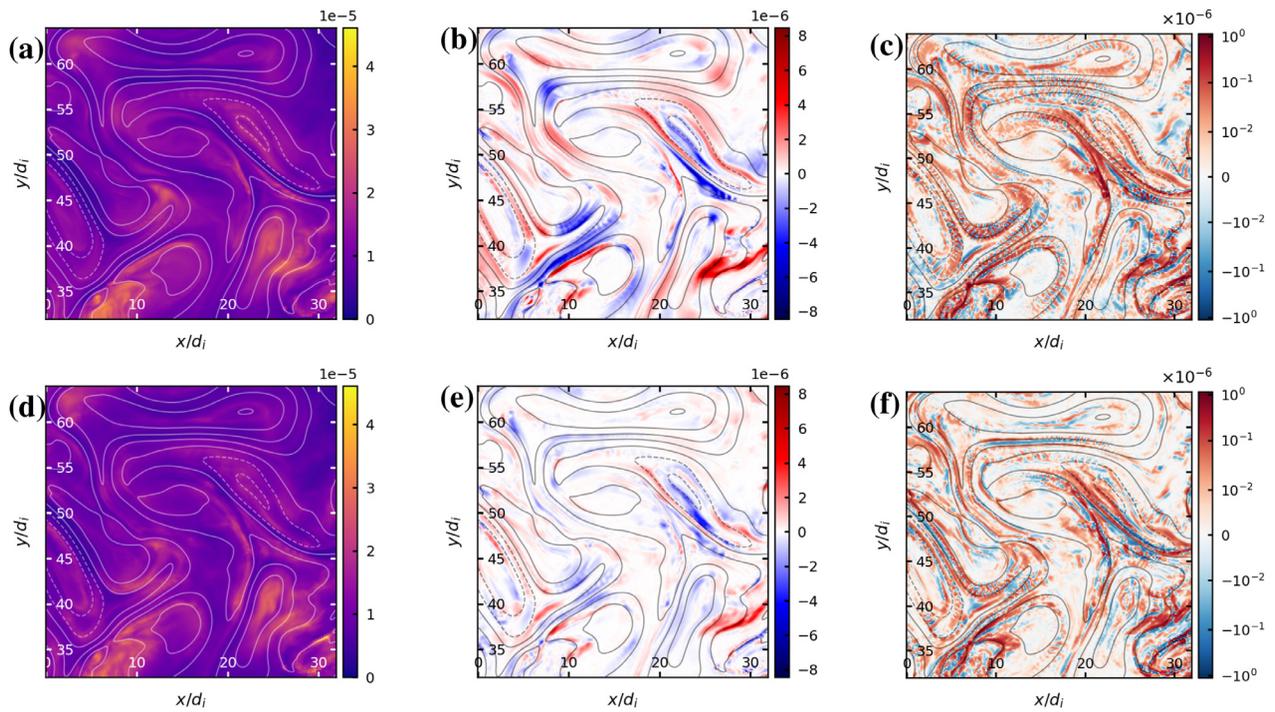


FIG. 8. Evaluation of the FCNN closures on a frame $t = 525 \omega_{pi}^{-1}$, with run A serving as a test set (see Table II). (a) Ground truth P_{xx} . (b) Ground truth P_{xy} . (c) Ground truth PiD . (d) FCNN predicted P_{xx} . (e) FCNN predicted P_{xy} . (f) FCNN predicted PiD . Each panel is equipped with the corresponding colormap. To provide a reference, we add contours of A_z , which is equivalent to the flux function in 2D.

Fig. 8(e) at that location reveals a similar pattern of P_{xy} , albeit less intense. In general, P_{xy} served by FCNN appears less intense and a bit more patchy, but even some small-scale structures coincide. We turn toward comparison between the incompressible part of pressure-strain

PiD in Fig. 8(c) and the predicted in Fig. 8(f). We see that in both cases, PiD tends to have large positive values on the separatrices. Comparison between other structures is also consistent, with some exceptions. For instance, the X point at $15d_i, 60d_i$ has a mismatch in the polarity of PiD .

We also study the Galilean invariance of the noE model in Appendix C.

We conclude the results section with the analog of Fig. 6(b) with the same analysis performed to obtain $\langle PiD | Q > Q^* \rangle$, applied to the dataset A and plotted in Fig. 9. The blue lines represent the ground truth, showing consistency with Fig. 6(b) in regard to Q_ω and Q_D conditionals, but showing that the A dataset has stronger PiD conditionals to Q_J and diverges even more from the result of Yang *et al.*⁶⁷ The prediction of FCNN matches the $\langle PiD | Q > Q^* \rangle$ behavior qualitatively but shows a different large Q threshold tail for all three conditionals. We note, however, that for large $Q > 4$ this statistic is not reliable because the larger the Q , the fewer pixels are actually used to compute these quantities of interest.

IV. DISCUSSION

In this manuscript, we introduce a new non-local neural closure for the electrons in the turbulent magnetosheath. This closure is obtained by training a fully convolutional neural network (FCNN) on the output of a fully kinetic particle-in-cell simulation of the energy-conserving code ECsim.¹¹ We submit this closure to multiple statistical evaluations on a hold-out test set, which comes from two separate numerical simulations. One test set comes from run B1 (tabulated

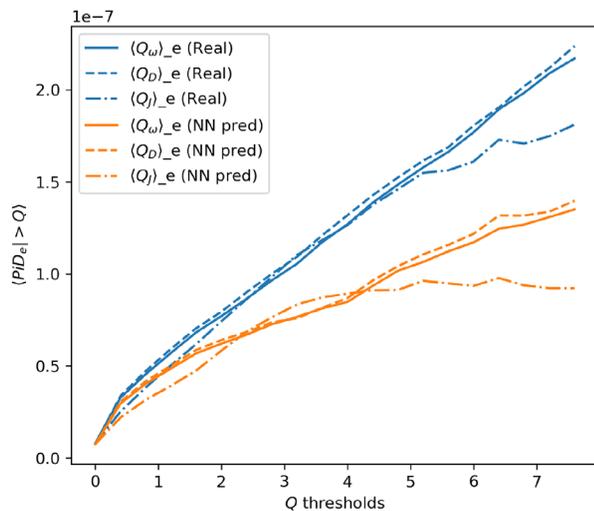


FIG. 9. Average incompressible pressure-strain PiD , conditional to values of quantities Q_ω , Q_D , and Q_J above a certain threshold [see Eqs. (18) and (19)]—dashed line, (20)—solid line, and (21)—dotted-dashed line. This quantity is computed for the ground truth (blue curves) and neural network prediction (orange curves).

in Table II) with the exact same ECsim parameters but a different random initialization. The other test set comes from run A, which is a simulation performed earlier⁵⁸ with a much larger number of particles and slightly different physical parameters (see Table I). Because of this, performance on run A can be considered a generalization test.

The closure we obtain is a five-moment closure for the electron pressure tensor as a function of density, electron velocity, magnetic field, and electric field. It can be considered a neural generalization of trapped/passing particle closure $p = p(n, B)$ ³¹ [see Eq. (9)], which is a symbolic expression and is itself a generalization of CGL.³⁰ We also compare it to architecture similar to the one used by Laperre *et al.*,⁴⁶ which consists of a multi-layer perceptron (MLP), a more traditional neural network that we apply locally, pointwise. In contrast, FCNN is applied to patches and can thus be considered a non-local closure. We benchmark the performance of FCNN, MLP, and more traditional symbolic anisotropic p_{\parallel} and p_{\perp} closures.³¹ We also attempted to learn a 10-moment closure for the heat flux using an FCNN architecture, but we found the performance to be relatively poor, which could be attributed to the limited data available. In general, we found that the 5-moment pressure tensor closure performed significantly worse when trained on a single simulation than when trained on four simulations.

In general, for the diagonal elements, we find that MLP provides results identical to symbolic closure,³¹ but FCNN identifies many of the mesoscale structures more accurately (see Sec. III A), while introducing some vapor-like noise. Furthermore, MLP does not capture well the quantities related to off-diagonal elements of the pressure tensor, such as agyrotropy (see Sec. III B). MLP underestimates the distributional spread in anisotropies, as well as derived quantities such as pressure-strain (Sec. III C), which is quite important for estimating the energy budget and energy channels in turbulence. In contrast, FCNN closure captures the important structures in the off-diagonal pressure tensor, albeit not perfectly. This is confirmed by lower values of determination score $R^2 \sim 0.4$ compared to diagonal components $R^2 \sim 0.75$. Nevertheless, the derived quantities from the FCNN-computed pressure tensor, such as the spatial distribution of pressure-strain and scale-to-scale filtered incompressible pressure-strain, appear quite similar to the results from direct numerical simulation (DNS). In particular, the budget of mean incompressible pressure-strain PiD , conditional on coherent structures defined according to the approach of Yang *et al.*,⁶⁷ reveals agreement between the actual and predicted values by the FCNN. The mean scale-to-scale filtered distribution of pressure-strain over the spatial scales, down to the smallest scales, is also consistent when comparing FCNN to the ground truth, while MLP tends to underestimate it.

To gain insight into the importance of pressure-strain features, such as the electric field, for instance, we train a selected set of pressure tensor models with different inputs in Sec. III D. Of particular interest are the following models: *default*, which is the standard model that takes as input all the lower order moments, and *noE*, which omits the electric field. These models are evaluated on two sets, A and B1, corresponding to very large and small number of particles per cell, respectively (see Table I). We find that *default* drastically under-performs on the generalization dataset A, while the score for *noE* is essentially unchanged. This is confirmed when carefully inspecting the spatial structures of P_{xx} , P_{xy} , and PiD associated with current sheets in the simulation. This leads to the conclusion that the electric field is not a

reliable predictor of the closure, at least when the closure is trained on more noisy simulations in set B. Note that, as mentioned in Sec. II E, we attempted to reduce this noise by applying Gaussian filters. For the discussion on generalization with respect to Galilean boosts, see Appendix C.

The fact that *noE* is able to generalize to the simulation with a different number of particles is reassuring, since certain small-scale structures are not consistent across the runs, and indicates that when the input parameters are chosen properly, the closure appears robust. We emphasize that this was achieved for only one set of physical parameters, namely, $\delta B/B \sim 0.6 - 0.7$, $\beta_i \sim 5.3 - 5.7$, and $\beta_e \sim 1.3 - 1.4$, consistent with magnetosheath conditions, and in 2D. In the future, we plan to train on a broader set of conditions, consisting of a parameter sweep over these quantities, to find a closure that interpolates between these regimes and extend our method to 3D. Generally, such parameter changes entail *out-of-distribution shifts* (OOD) in the relevant quantities, such as density or pressure. This problem is usually treated with *transfer learning*. This implies taking a pre-trained base network and fine-tuning some of the network's layers in response to distribution shifts in the new physical conditions. This approach has shown promise in large eddy simulations (LESs) in the work of Subel *et al.*⁷³ across several cases: adapting to changes in the forcing wavelength and increasing the Reynolds number. We propose undertaking such work in plasma physics in the future. The problem of introducing a neural network into a parametrization of a physical process is a challenging one and requires a series of adaptations, including training on multiple-step roll-outs.^{73,74} This implies wrapping a numerical solver inside the loss function of the optimization algorithm, a method also referred to as *online (a posteriori) training and testing*. Recently, there have been developments on promoting globally stable data-driven models, which can be achieved via modifications of the objective function⁷⁵ or via enforcing hyperbolicity.⁷⁶

Additional steps can be taken in the future to improve the quality of the pressure tensor closure. It is clear that with more training data, better results may be achieved; however, this implies computational costs associated with generating such data, especially since some runs must be reserved for validation. In principle, many groups worldwide possess particle-in-cell or Vlasov simulation data that could be useful for training such closures. Thus, one could conceive of deeper neural network architectures trained on the wealth of simulation data, mirroring the work in meteorology where neural surrogate models were trained on 40 years of reanalysis data, rivaling numerical weather prediction models.^{44,77} The latter is, of course, possible thanks to the existence of high-quality data obtained on assimilating observations into models,⁷⁸ a dataset of such quality we do not possess in the context of space plasmas. Thus, from this angle, developing data-driven models that outperform existing high-fidelity numerical simulations such as ECsim^{11,57} does not appear to be feasible in the near future; however, training neural surrogates that are more efficient than conventional methods can still be achieved. In particular, the closures obtained in this way can be embedded in reduced order models (ROMs) such as fluid⁷⁹ and hybrid kinetic models^{17,18} that are more computationally tractable than fully kinetic simulations and with the appropriate closure can reproduce the outputs of such fully kinetic simulations.^{32,38-40} Similar closures allow the study of kinetic processes such as kinetic Alfen wave (KAW) turbulence on a larger span of inertial range.^{80,81}

The goal of data-driven closure is to extend the validity of ROMs, which, thanks to their efficiency, can be simulated more frequently and across a broader range of parameters than high-fidelity models.

We would like to emphasize a few other avenues that could guide data-driven closure development. First, in this manuscript, we have experimented with rather traditional architectures, such as MLP and FCNN, which leave room for more modern AI models, including neural operators, such as Fourier neural operators (FNOs).^{82,83} FNO has already been applied for neural surrogate modeling of a plasma fusion device.⁸⁴ Furthermore, we have relied solely on the standard loss function, mean squared error (MSE), and have not exploited *soft constraints*,⁸⁵ additional physics-based constraints that can enhance the physical fidelity of the learned representation.

Another promising line of research for obtaining closure relations is equation discovery,⁵¹ which is a collection of methods that extract equations from data using symbolic or sparse regression. In sparse regression, a library of preselected expressions is fitted.^{52–55} These methods have also been applied in conjunction with data augmentation,⁸⁶ such as applying Lorenz/Galilean boosts that enforce such invariance and improve the fidelity of the learned models. We would like to emphasize that methods such as physics-informed sparse regression,⁸⁷ genetic programming, and pre-trained transformers⁵⁶ warrant more attention in plasma physics with regard to these types of problems. Naturally, as is the case with other forms of machine learning, these methods are prone to overfitting when presented with partial data and high expressivity (complexity of the expressions that can be fitted by the method). This is where intuition regarding physics-based closures^{62,63} can be very useful in restricting the set of possibilities *a priori*. We firmly believe that progress in this field is possible by a careful combination of machine learning, high-performance computing, and theoretical considerations.

V. CONCLUSION

This work is the first application of non-local neural closure for the electron pressure tensor, achieved via a fully convolutional neural network (FCNN). Using a combination of statistical and physical fidelity diagnostics, such as pressure-strain and agyrotropy, we have demonstrated the generalization of this new closure from noisy (fewer particles per cell) particle-in-cell (PIC) simulations to more accurate (higher particle counts per cell) simulations. Pressure-strain diagnostics indicate that the closure accurately captures overall energy channels and certain local characteristics near the X-point of the reconnection site. This is promising, as we run PIC simulations for training data generation; however, with a higher number of particles per cell, simulations become prohibitively expensive to run in large quantities. Crucially, we demonstrate that FCNN significantly outperforms known closure relations, such as the previously used multi-layer perceptron (MLP) or other double adiabatic-type models. We have addressed this problem in the context of Earth's magnetosheath decaying turbulence simulations, considering a specific set of physical parameters associated with large ion β and moderate electron β . Future work will involve extending the validity of this closure to a broader set of parameters, 3D geometry, and coupling it to reduced order models (ROMs), such as two-fluid and hybrid kinetic simulations. It will contribute to the development of efficient multi-scale models capable of probing larger domains of magnetospheric physics while accurately representing small-scale physics.

ACKNOWLEDGMENTS

G.M. dedicates this work to the memory of Giovanni Lapenta, whose guidance, insight, and goodwill were essential.

The authors would like to thank Thierry Passot, Pierre-Henri, Silvio Cerri, Francesco Carella, Rony Keppens, Stefaan Poedts, and Maria Elena Innocenti for their valuable discussions and support.

G. Miloshevich has received funding from the European Union's Horizon research and innovation program under the Marie Skłodowska-Curie (Grant No. 101148539). N. F. Oliveira Lopes and P. Dazzi would like to acknowledge the Research Foundation—Flanders (FWO), HELIOSKILL project (Grant No. G0B9923N).

The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation—Flanders (FWO) and the Flemish Government.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

G. Miloshevich: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (lead); Resources (lead); Software (lead); Supervision (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **L. Vranckx:** Data curation (equal); Formal analysis (equal); Investigation (equal); Validation (equal); Writing – review & editing (equal). **F. N. de Oliveira Lopes:** Software (supporting); Validation (equal); Visualization (equal); Writing – review & editing (supporting). **P. Dazzi:** Investigation (supporting); Validation (supporting); Writing – review & editing (equal). **G. Arrò:** Conceptualization (supporting); Data curation (equal); Methodology (supporting); Supervision (supporting); Validation (supporting); Writing – review & editing (supporting). **G. Lapenta:** Conceptualization (supporting); Data curation (supporting); Funding acquisition (supporting); Methodology (supporting); Project administration (supporting); Supervision (supporting).

DATA AVAILABILITY

The test datasets A and B1 used as test sets in this study are available in the Zenodo repository at <https://zenodo.org/records/17882782>. The code is available at <https://github.com/georgemilosh/closure.git>.

APPENDIX A: SCALE-TO-SCALE ANALYSIS

To investigate the behavior of energy conversion channels at different length scales, we employ the scale-filtering/coarse-graining method, widely used to analyze both magnetohydrodynamic^{88–90} and plasma^{58,67,68} turbulence. We introduce a general filtering operation

$$\bar{f}_\ell^s(\mathbf{x}, t) = \int d^d r G_\ell(\mathbf{r}) f_s(\mathbf{x} + \mathbf{r}, t), \quad (\text{A1})$$

where, for the remainder of the manuscript, we will only use the boxcar filter following Matthaeus.⁹¹ In addition, we will need the “ Favre-filtered ” (density-weighted filtered) quantities

$$\tilde{f}_\ell^s = \overline{\rho f_\ell^s} / \bar{\rho}_\ell^s. \quad (\text{A2})$$

We present a scale-to-scale filtered version of Eq. (17) as follows, where we have removed the spatial transport terms by performing spatial box averaging $\langle \cdot \rangle$:

$$\partial_t \langle \tilde{E}_s^f \rangle = -\langle \Pi_s^{VV} \rangle - \langle \Phi_s^{VT} \rangle - \langle \Lambda_s^{Vb} \rangle, \quad (\text{A3a})$$

$$\partial_t \langle \tilde{E}^m \rangle = -\sum_s \langle \Pi_s^{bb} \rangle + \sum_s \langle \Lambda_s^{Vb} \rangle. \quad (\text{A3b})$$

TABLE IV. Ablation study on datasplit 2 (see Table II): comparison between FCNN trained on different inputs. Model referred to as “default” consists of $(n, \mathbf{v}_e, \mathbf{E}, \mathbf{B})$ inputs, “noE” corresponds to $(n, \mathbf{v}_e, \mathbf{B})$, “Jtot” corresponds to $(n, \mathbf{J}, \mathbf{E}, \mathbf{B})$, and “JtotnoE” corresponds to $(n, \mathbf{J}, \mathbf{B})$.

Model	Default	noE	Jtot	JtotnoE
P_{\parallel}	0.78	0.79	0.78	0.81
P_{\perp}	0.84	0.82	0.83	0.84
P_{xx}	0.81	0.80	0.80	0.82
P_{yy}	0.82	0.80	0.81	0.83
P_{zz}	0.77	0.77	0.77	0.80
P_{xy}	0.42	0.37	0.42	0.37
P_{xz}	0.37	0.33	0.38	0.35
P_{yz}	0.37	0.34	0.38	0.36

Spatial averaging symbol will be omitted for the remainder of the manuscript. Here, the *filtered fluid flow energy* is given by

$$\tilde{E}_s^f = \frac{1}{2} \bar{\rho}_s \tilde{V}_s^2, \quad (\text{A4})$$

and the *filtered electromagnetic energy* is

$$\bar{E}^m = \frac{\bar{\mathbf{B}}^2 + \bar{\mathbf{E}}^2}{8\pi}. \quad (\text{A5})$$

The *sub-grid-scale (SGS) flux of fluid flow energy across scales* due to nonlinearities is

$$\Pi_s^{VV} = -\left(\bar{\rho}_s \tilde{\tau}_s^V \cdot \nabla\right) \cdot \tilde{\mathbf{V}}_s - \frac{q_s}{c} \bar{n}_s \tilde{\tau}_s^b \cdot \tilde{\mathbf{V}}_s, \quad (\text{A6})$$

where

$$\tilde{\tau}_s^V = \widetilde{\mathbf{V}_s \mathbf{V}_s} - \tilde{V}_s \tilde{V}_s, \quad \tilde{\tau}_s^b = \mathbf{V}_s \times \widetilde{\mathbf{B}} - \tilde{V}_s \times \tilde{\mathbf{B}}. \quad (\text{A7})$$

The *SGS flux of electromagnetic energy across scales* due to nonlinearities is

$$\Pi_s^{bb} = -q_s \bar{n}_s \tilde{\tau}_s^e \cdot \tilde{\mathbf{V}}_s, \quad \text{where } \tilde{\tau}_s^e = \tilde{\mathbf{E}} - \bar{\mathbf{E}}. \quad (\text{A8})$$

The *rate of conversion of flow energy into internal energy* is

$$\Phi_s^{VT} = -(\bar{\mathbf{P}}_s \cdot \nabla) \cdot \tilde{\mathbf{V}}_s. \quad (\text{A9})$$

The *rate of conversion of fluid flow energy into electromagnetic energy* is

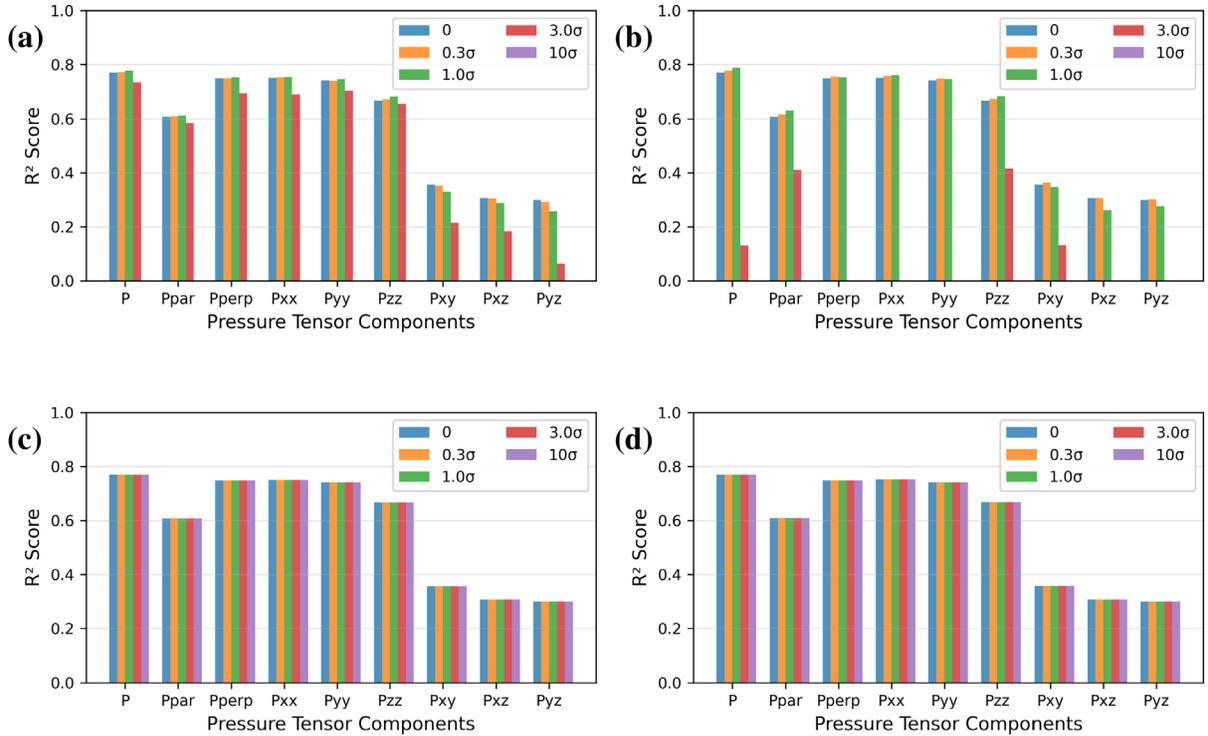


FIG. 10. Galilean invariance test over four different values of boosts corresponding to $v/c = \{0, 0.005, 0.02, 0.05\}$. (a) Applying $\mathbf{v} = v \hat{x}$. (b) Applying $\mathbf{v} = v \hat{z}$. (c) Applying $\mathbf{v} = v \hat{x}$ and subtracting the mean v_x from the input to the neural network (d), applying $\mathbf{v} = v \hat{z}$ and subtracting the mean v_z from the input to the neural network.

$$\Lambda_s^{vb} = -q_s \bar{n}_s \tilde{\mathbf{E}} \cdot \tilde{\mathbf{V}}_s. \quad (\text{A10})$$

APPENDIX B: DATASPLIT 2

In Table IV, we present results of study similar to Table III and Fig. 7(a) but applied to a different datasplit, see Table II. It appears that R^2 score for both diagonal and off-diagonal pressure tensor is consistent, which bolsters the robustness of the study.

APPENDIX C: GALILEAN INVARIANCE

In what follows, we apply the Galilean boosts $\mathbf{x}' = \mathbf{x} - \mathbf{v}t$, which are also associated with the transformation of the electromagnetic field that under $\mathbf{v} \ll c$

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B}, \quad (\text{C1})$$

and

$$\mathbf{B}' = \mathbf{B} - \frac{1}{c^2} \mathbf{v} \times \mathbf{E}. \quad (\text{C2})$$

The primes denote quantities in the boosted reference frame. The values are provided as inputs. These transformations are applied to the neural network's input to predict pressure. We plot the determination scores for a range of values of boosts $v/c = \{0, 0.005, 0.02, 0.05, 0.2\}$ in Fig. 10. The values are chosen to correspond to fractions of variance of the corresponding component of the velocity $\sigma_x := \sqrt{\langle v_x^2 \rangle}$ in Fig. 10(a) and $\sigma_z := \sqrt{\langle v_z^2 \rangle}$ in Fig. 10(b). From Fig. 10(a), we see that the network is capable of reconstructing the diagonal pressure tensor components up to 3σ . From Fig. 10(b), we see that performance drops to 1σ , above which the network fails. This is because as we increase \mathbf{v} , the network is supplied with a larger distribution of velocity that it has not seen during training. One way to address this problem is to augment the dataset by applying random Galilean boosts during training. This is similar to the recent work⁸⁶ that embeds Lorentz invariance via data augmentation. However, for our purpose, we can show that this is not necessary since we can simply augment the pressure model (12) using the following anzats:

$$\mathbf{P} = \mathbf{P}_0(n, \mathbf{V}_e - \langle \mathbf{V}_e \rangle, \mathbf{E}, \mathbf{B}), \quad (\text{C3})$$

where $\langle \rangle$ operation implies spatial averaging. This is meant to undo the main contribution of the Galilean boost, and no additional training is needed, since during training $\langle \mathbf{V}_e \rangle \approx 0$. This trivial operation yields updated figures for both the v_x [Fig. 10(c)] and v_z boosts [Fig. 10(d)], and we can see that the performance of closure is independent of the boost value because the changes in B field [Eq. (C2)], which are also inputs to Eq. (C3), are subdominant.

REFERENCES

- ¹J. E. Stawarz, J. P. Eastwood, T. D. Phan, I. L. Gingell, P. S. Pyakurel, M. A. Shay, S. L. Robertson, C. T. Russell, and O. Le Contel, "Turbulence-driven magnetic reconnection and the magnetic correlation length: Observations from magnetospheric multiscale in Earth's magnetosheath," *Phys. Plasmas* **29**, 012302 (2022).
- ²A. Retinò, Y. Khotyaintsev, O. Le Contel, M. F. Marcucci, F. Plaschke, A. Vaivads, V. Angelopoulos, P. Blasi, J. Burch, J. De Keyser, M. Dunlop, L. Dai, J. Eastwood, H. Fu, S. Haaland, M. Hoshino, A. Johlander, L. Kepko, H. Kucharek, G. Lapenta, B. Lavraud, O. Malandraki, W. Matthaeus, K. McWilliams, A. Petrukovich, J.-L. Pinçon, Y. Saito, L. Sorriso-Valvo, R. Vainio, and R. Wimmer-Schweingruber, "Particle energization in space plasmas: Towards a multi-point, multi-scale plasma observatory," *Exp. Astron.* **54**, 427–471 (2022).
- ³THIRA, see https://www.fema.gov/sites/default/files/2020-06/fema_national-thira-overview-methodology_2019_0.pdf for "2019 National Threat and Hazard Identification and Risk Assessment (THIRA): Overview and methodology" (2019).
- ⁴A. Lazarian, G. Eyink, E. Vishniac, and G. Kowal, "Turbulent reconnection and its implications," *Philos. Trans. R. Soc., A* **373**, 20140144 (2015).
- ⁵A. Lazarian, G. L. Eyink, A. Jafari, G. Kowal, H. Li, S. Xu, and E. T. Vishniac, "3D turbulent reconnection: Theory, tests, and astrophysical implications," *Phys. Plasmas* **27**, 012305 (2020).
- ⁶J. E. Stawarz, P. A. Muñoz, N. Besho, R. Bandyopadhyay, T. K. M. Nakamura, S. Eriksson, D. B. Graham, J. Büchner, A. Chasapis, J. F. Drake, M. A. Shay, R. E. Ergun, H. Hasegawa, Y. V. Khotyaintsev, M. Swisdak, and F. D. Wilder, "The interplay between collisionless magnetic reconnection and turbulence," *Space Sci. Rev.* **220**, 90 (2024).
- ⁷R. Nakamura, J. L. Burch, J. Birn, L.-J. Chen, D. B. Graham, F. Guo, K.-J. Hwang, H. Ji, Y. V. Khotyaintsev, Y.-H. Liu, M. Oka, D. Payne, M. I. Sitnov, M. Swisdak, S. Zenitani, J. F. Drake, S. A. Fuselier, K. J. Genestreti, D. J. Gershman, H. Hasegawa, M. Hoshino, C. Norgren, M. A. Shay, J. R. Shuster, and J. E. Stawarz, "Outstanding questions and future research on magnetic reconnection," *Space Sci. Rev.* **221**, 17 (2025).
- ⁸Y. V. Khotyaintsev, D. B. Graham, C. Norgren, and A. Vaivads, "Collisionless magnetic reconnection and waves: Progress review," *Front. Astron. Space Sci.* **6**, 70 (2019).
- ⁹G. Arrò, F. Pucci, F. Califano, M. E. Innocenti, and G. Lapenta, "Generation of subion scale magnetic holes from electron shear flow instabilities in plasma turbulence," *Astrophys. J.* **958**, 11 (2023).
- ¹⁰J. Espinoza-Troni, G. Arrò, F. A. Asenjo, and P. S. Moya, "Electron scale magnetic holes generation driven by Whistler-to-Bernstein mode conversion in fully kinetic plasma turbulence," [arXiv:2501.09651](https://arxiv.org/abs/2501.09651) (2025).
- ¹¹G. Lapenta, "Exactly energy conserving semi-implicit particle in cell formulation," *J. Comput. Phys.* **334**, 349–366 (2017).
- ¹²R. A. Fonseca, L. O. Silva, F. S. Tsung, V. K. Decyk, W. Lu, C. Ren, W. B. Mori, S. Deng, S. Lee, T. Katsouleas, and J. C. Adam, "OSIRIS: A three-dimensional, fully relativistic particle in cell code for modeling plasma based accelerators," in *Proceedings of the Computational Science — ICCS 2002*, edited by P. M. A. Sloot, A. G. Hoekstra, C. J. K. Tan, and J. J. Dongarra (Springer, Berlin, Heidelberg, 2002), pp. 342–351.
- ¹³R. A. Fonseca, S. F. Martins, L. O. Silva, J. W. Tonge, F. S. Tsung, and W. B. Mori, "One-to-one direct modeling of experiments and astrophysical scenarios: Pushing the envelope on kinetic plasma simulations," *Plasma Phys. Controlled Fusion* **50**, 124034 (2008).
- ¹⁴K. Germaschewski, W. Fox, S. Abbott, N. Ahmadi, K. Maynard, L. Wang, H. Ruhl, and A. Bhattacharjee, "The Plasma Simulation Code: A modern particle-in-cell code with patch-based load-balancing," *J. Comput. Phys.* **318**, 305–326 (2016).
- ¹⁵F. Allmann-Rahn, S. Lautenbach, M. Deisenhofer, and R. Grauer, "The muphyII code: Multiphysics plasma simulation on large HPC systems," *Comput. Phys. Commun.* **296**, 109064 (2024).
- ¹⁶F. Valentini, P. Trávníček, F. Califano, P. Hellinger, and A. Mangeney, "A hybrid-Vlasov model based on the current advance method for the simulation of collisionless magnetized plasma," *J. Comput. Phys.* **225**, 753–770 (2007).
- ¹⁷M. Palmroth, U. Ganse, Y. Pfau-Kempf, M. Battarbee, L. Turc, T. Brito, M. Grandin, S. Hoilijoki, A. Sandroos, and S. von Althaus, "Vlasov methods in space physics and astrophysics," *Living Rev. Comput. Astrophys.* **4**(1), 1 (2018).
- ¹⁸E. Behar, S. Fatemi, P. Henri, and M. Holmström, "Menura: A code for simulating the interaction between a turbulent solar wind and solar system bodies," *Ann. Geophys.* **40**, 281–297 (2022).
- ¹⁹C. Dong, L. Wang, A. Hakim, A. Bhattacharjee, J. A. Slavin, G. A. DiBraccio, and K. Germaschewski, "Global ten-moment multifluid simulations of the solar wind interaction with mercury: From the planetary conducting core to the dynamic magnetosphere," *Geophys. Res. Lett.* **46**, 11584–11596, <https://doi.org/10.1029/2019GL083180> (2019).

- ²⁰X. Wang, Y. Chen, and G. Tóth, “Global magnetohydrodynamic magnetosphere simulation with an adaptively embedded particle-in-cell model,” *J. Geophys. Res. Space Phys.* **127**, e2021JA030091, <https://doi.org/10.1029/2021JA030091> (2022).
- ²¹J. Birn, J. F. Drake, M. A. Shay, B. N. Rogers, R. E. Denton, M. Hesse, M. Kuznetsova, Z. W. Ma, A. Bhattacharjee, A. Otto, and P. L. Pritchett, “Geospace environmental modeling (GEM) magnetic reconnection challenge,” *J. Geophys. Res. Space Phys.* **106**, 3715–3719, <https://doi.org/10.1029/1999JA900449> (2001).
- ²²M. Hesse and P. A. Cassak, “Magnetic reconnection in the space sciences: Past, present, and future,” *J. Geophys. Res. Space Phys.* **125**, e2018JA025935, <https://doi.org/10.1029/2018JA025935> (2020).
- ²³G. Lapenta, D. Schriver, R. J. Walker, J. Berchem, N. F. Echterling, M. El Alaoui, and P. Travnicek, “Do we need to consider electrons’ kinetic effects to properly model a planetary magnetosphere: The case of mercury,” *J. Geophys. Res. Space Phys.* **127**, e2021JA030241, <https://doi.org/10.1029/2021JA030241> (2022).
- ²⁴P. M. Trávníček, D. Schriver, P. Hellinger, D. Herčík, B. J. Anderson, M. Sarantos, and J. A. Slavin, “Mercury’s magnetosphere–solar wind interaction for northward and southward interplanetary magnetic field: Hybrid simulation results,” *Icarus* **209**, 11–22 (2010).
- ²⁵L. K. S. Daldorff, G. Tóth, T. I. Gombosi, G. Lapenta, J. Amaya, S. Markidis, and J. U. Brackbill, “Two-way coupling of a global Hall magnetohydrodynamics model with a local implicit particle-in-cell model,” *J. Comput. Phys.* **268**, 236–254 (2014).
- ²⁶G. Tóth, X. Jia, S. Markidis, I. B. Peng, Y. Chen, L. K. S. Daldorff, V. M. Tennishev, D. Borovikov, J. D. Haiducek, T. I. Gombosi, A. Gloer, and J. C. Dorelli, “Extended magnetohydrodynamics with embedded particle-in-cell simulation of Ganymede’s magnetosphere,” *J. Geophys. Res. Space Phys.* **121**, 1273–1293, <https://doi.org/10.1002/2015JA021997> (2016).
- ²⁷X. Wang, Y. Chen, and G. Tóth, “Simulation of magnetospheric sawtooth oscillations: The role of kinetic reconnection in the magnetotail,” *Geophys. Res. Lett.* **49**, e2022GL099638, <https://doi.org/10.1029/2022GL099638> (2022).
- ²⁸Y. Chen, G. Tóth, H. Zhou, and X. Wang, “FLEKS: A flexible particle-in-cell code for multi-scale plasma simulations,” *Comput. Phys. Commun.* **287**, 108714 (2023).
- ²⁹D. Li, Y. Chen, C. Dong, L. Wang, and G. Toth, “Numerical study of magnetic island coalescence using magnetohydrodynamics with adaptively embedded particle-in-cell model,” *AIP Adv.* **13**, 015126 (2023).
- ³⁰G. F. Chew, M. L. Goldberger, F. E. Low, and S. Chandrasekhar, “The Boltzmann equation and the one-fluid hydromagnetic equations in the absence of particle collisions,” *Proc. R. Soc. London, Ser. A* **236**, 112–118 (1956).
- ³¹A. Le, J. Egedal, W. Daughton, W. Fox, and N. Katz, “Equations of state for collisionless guide-field reconnection,” *Phys. Rev. Lett.* **102**, 085001 (2009).
- ³²O. Ohia, J. Egedal, V. S. Lukin, W. Daughton, and A. Le, “Demonstration of anisotropic fluid closure capturing the kinetic structure of magnetic reconnection,” *Phys. Rev. Lett.* **109**, 115004 (2012).
- ³³B. A. Wetheron, J. Egedal, A. Lê, and W. Daughton, “Validation of anisotropic electron fluid closure through in situ spacecraft observations of magnetic reconnection,” *Geophys. Res. Lett.* **46**, 6223–6229, <https://doi.org/10.1029/2019GL083119> (2019).
- ³⁴J. L. Burch, R. B. Torbert, T. D. Phan, L.-J. Chen, T. E. Moore, R. E. Ergun, J. P. Eastwood, D. J. Gershman, P. A. Cassak, M. R. Argall, S. Wang, M. Hesse, C. J. Pollock, B. L. Giles, R. Nakamura, B. H. Mauk, S. A. Fuselier, C. T. Russell, R. J. Strangeway, J. F. Drake, M. A. Shay, Y. V. Khotyaintsev, P.-A. Lindqvist, G. Marklund, F. D. Wilder, D. T. Young, K. Torkar, J. Goldstein, J. C. Dorelli, L. A. Avanov, M. Oka, D. N. Baker, A. N. Jaynes, K. A. Goodrich, I. J. Cohen, D. L. Turner, J. F. Fennell, J. B. Blake, J. Clemmons, M. Goldman, D. Newman, S. M. Petrinec, K. J. Trattner, B. Lavraud, P. H. Reiff, W. Baumjohann, W. Magnes, M. Steller, W. Lewis, Y. Saito, V. Coffey, and M. Chandler, “Electron-scale measurements of magnetic reconnection in space,” *Science* **352**, aaf2939 (2016).
- ³⁵M. Hesse, M. Kuznetsova, and J. Birn, “The role of electron heat flux in guide-field magnetic reconnection,” *Phys. Plasmas* **11**, 5387–5397 (2004).
- ³⁶G. W. Hammett and F. W. Perkins, “Fluid moment models for Landau damping with application to the ion-temperature-gradient instability,” *Phys. Rev. Lett.* **64**, 3019–3022 (1990).
- ³⁷T. Passot and P. L. Sulem, “Collisionless magnetohydrodynamics with gyrokinetic effects,” *Phys. Plasmas* **14**, 082502 (2007).
- ³⁸F. Finelli, S. S. Cerri, F. Califano, F. Pucci, D. Laveder, G. Lapenta, and T. Passot, “Bridging hybrid- and full-kinetic models with Landau-fluid electrons—I. 2D magnetic reconnection,” *Astron. Astrophys.* **653**, A156 (2021).
- ³⁹L. Wang, A. H. Hakim, A. Bhattacharjee, and K. Germaschewski, “Comparison of multi-fluid moment models with particle-in-cell simulations of collisionless magnetic reconnection,” *Phys. Plasmas* **22**, 012108 (2015).
- ⁴⁰J. Ng, A. Hakim, A. Bhattacharjee, A. Stanier, and W. Daughton, “Simulations of anti-parallel reconnection using a nonlocal heat flux closure,” *Phys. Plasmas* **24**, 082112 (2017).
- ⁴¹A. Stanier, W. Daughton, L. Chacón, H. Karimabadi, J. Ng, Y.-M. Huang, A. Hakim, and A. Bhattacharjee, “Role of ion kinetic physics in the interaction of magnetic flux ropes,” *Phys. Rev. Lett.* **115**, 175004 (2015).
- ⁴²R. Vinuesa and S. L. Brunton, “Enhancing computational fluid dynamics with machine learning,” *Nat. Comput. Sci.* **2**(6), 358–366 (2022).
- ⁴³R. Anirudh, R. Archibald, M. S. Asif, M. M. Becker, S. Benkadda, P.-T. Bremer, R. H. S. Budé, C. S. Chang, L. Chen, R. M. Churchill, J. Citrin, J. A. Gaffney, A. Gainaru, W. Gekelman, T. Gibbs, S. Hamaguchi, C. Hill, K. Humbird, S. Jalas, S. Kawaguchi, G.-H. Kim, M. Kirchen, S. Klasky, J. L. Kline, K. Krushelnick, B. Kustowski, G. Lapenta, W. Li, T. Ma, N. J. Mason, A. Mesbah, C. Michoski, T. Munson, I. Murakami, H. N. Najm, K. E. J. Olofsson, S. Park, J. L. Peterson, M. Probst, D. Pugmire, B. Sammulu, K. Sawlani, A. Scheinker, D. P. Schissel, R. J. Shaloo, J. Shinagawa, J. Seong, B. K. Spears, J. Tennyson, J. Thiagarajan, C. M. Ticoş, J. Trieschmann, J. van Dijk, B. Van Essen, P. Ventzek, H. Wang, J. T. L. Wang, Z. Wang, K. Wende, X. Xu, H. Yamada, T. Yokoyama, and X. Zhang, “2022 review of data-driven plasma science,” [arXiv:2205.15832](https://arxiv.org/abs/2205.15832) (2022).
- ⁴⁴D. Kochkov, J. Yuval, I. Langmore, P. Norgaard, J. Smith, G. Mooers, M. Klöwer, J. Lottes, S. Rasp, P. Düben, S. Hatfield, P. Battaglia, A. Sanchez-Gonzalez, M. Willson, M. P. Brenner, and S. Hoyer, “Neural general circulation models for weather and climate,” *Nature* **632**, 1060–1066 (2024).
- ⁴⁵P. Perezhogin, L. Zanna, and C. Fernandez-Granda, “Generative data-driven approaches for stochastic subgrid parameterizations in an idealized ocean model,” *J. Adv. Model. Earth Syst.* **15**, e2023MS003681 (2023).
- ⁴⁶B. Laperre, J. Amaya, S. Jamal, and G. Lapenta, “Identification of high order closure terms from fully kinetic simulations using machine learning,” *Phys. Plasmas* **29**, 032706 (2022).
- ⁴⁷C. Ma, B. Zhu, X.-Q. Xu, and W. Wang, “Machine learning surrogate models for Landau fluid closure,” *Phys. Plasmas* **27**, 042502 (2020).
- ⁴⁸R. Maulik, N. A. Garland, J. W. Burby, X.-Z. Tang, and P. Balaprakash, “Neural network representability of fully ionized plasma fluid model closures,” *Phys. Plasmas* **27**, 072106 (2020).
- ⁴⁹L. Bois, E. Franck, L. Navoret, and V. Vigon, “A neural network closure for the Euler-Poisson system based on kinetic simulations,” *Kinetic Related Models* **15**, 49–89 (2022).
- ⁵⁰Z. Huang, C. Dong, and L. Wang, “Machine-learning heat flux closure for multi-moment fluid modeling of nonlinear Landau damping,” *Proc. Nat. Acad. Sci.* **122**, e2419073122 (2025).
- ⁵¹G. Camps-Valls, A. Gerhardus, U. Ninad, G. Varando, G. Martius, E. Balaguer-Ballester, R. Vinuesa, E. Diaz, L. Zanna, and J. Runge, “Discovering causal relations and equations from data,” *Phys. Rep.* **1044**, 1–68 (2023).
- ⁵²E. P. Alves and F. Fiuza, “Data-driven discovery of reduced plasma physics models from fully kinetic simulations,” *Phys. Rev. Res.* **4**, 033192 (2022).
- ⁵³S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proc. Nat. Acad. Sci.* **113**, 3932–3937 (2016).
- ⁵⁴J. Donaghy and K. Germaschewski, “In search of a data-driven symbolic multi-fluid ten-moment model closure,” *J. Plasma Phys.* **89**, 895890105 (2023).
- ⁵⁵E. R. Ingelsten, M. C. McGrae-Menge, E. P. Alves, and I. Pusztai, “Data-driven discovery of a heat flux closure for electrostatic plasma phenomena,” *J. Plasma Phys.* **91**, E64 (2025).
- ⁵⁶M. Landajuela, C. S. Lee, J. Yang, R. Glatt, C. P. Santiago, I. Aravena, T. Mundhenk, G. Mulcahy, and B. K. Petersen, “A unified framework for deep symbolic regression,” in *36th Conference on Neural Information Processing Systems* (Curran Associates, Inc., 2022), Vol. 35, pp. 33985–33998, see https://proceedings.neurips.cc/paper_files/paper/2022/file/dbca58f35bddc6e4003b2dd80e42f838-Paper-Conference.pdf.

- ⁵⁷S. Markidis and G. Lapenta, “The energy conserving particle-in-cell method,” *J. Comput. Phys.* **230**, 7037–7052 (2011).
- ⁵⁸G. Arrò, F. Califano, and G. Lapenta, “Spectral properties and energy transfer at kinetic scales in collisionless plasma turbulence,” *Astron. Astrophys.* **668**, A33 (2022).
- ⁵⁹H. M. Abdelhamid, Y. Kawazura, and Z. Yoshida, “Hamiltonian formalism of extended magnetohydrodynamics,” *J. Phys. A: Math. Theor.* **48**, 235502 (2015).
- ⁶⁰M. Lingam, P. J. Morrison, and G. Miloshevich, “Remarkable connections between extended magnetohydrodynamics models,” *Phys. Plasmas* **22**, 072111 (2015).
- ⁶¹M. Lingam, G. Miloshevich, and P. J. Morrison, “Concomitant Hamiltonian and topological structures of extended magnetohydrodynamics,” *Phys. Lett. A* **380**, 2400–2406 (2016).
- ⁶²P. Hunana, A. Tenerani, G. P. Zank, M. L. Goldstein, G. M. Webb, E. Khomenko, M. Collados, P. S. Cally, L. Adhikari, and M. Velli, “An introductory guide to fluid models with anisotropic temperatures. Part 2. Kinetic theory, Padé approximants and Landau fluid closures,” *J. Plasma Phys.* **85**, 205850603 (2019).
- ⁶³P. Hunana, A. Tenerani, G. P. Zank, E. Khomenko, M. L. Goldstein, G. M. Webb, P. S. Cally, M. Collados, M. Velli, and L. Adhikari, “An introductory guide to fluid models with anisotropic temperatures. Part 1. CGL description and collisionless fluid hierarchy,” *J. Plasma Phys.* **85**, 205850602 (2019).
- ⁶⁴M. Swisdak, “Quantifying gyrotropy in magnetic reconnection,” *Geophys. Res. Lett.* **43**, 43–49, <https://doi.org/10.1002/2015GL066980> (2016).
- ⁶⁵F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain,” *Psychol. Rev.* **65**, 386–408 (1958).
- ⁶⁶D. Chicco, M. J. Warrens, and G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput. Sci.* **7**, e623 (2021).
- ⁶⁷Y. Yang, W. H. Matthaeus, T. N. Parashar, P. Wu, M. Wan, Y. Shi, S. Chen, V. Roytershteyn, and W. Daughton, “Energy transfer channels and turbulence cascade in Vlasov-Maxwell turbulence,” *Phys. Rev. E* **95**, 061201 (2017).
- ⁶⁸W. H. Matthaeus, Y. Yang, M. Wan, T. N. Parashar, R. Bandyopadhyay, A. Chasapis, O. Pezzi, and F. Valentini, “Pathways to dissipation in weakly collisional plasmas,” *Astrophys. J.* **891**, 101 (2020).
- ⁶⁹S. P. Gary and J. Wang, “Whistler instability: Electron anisotropy upper bound,” *J. Geophys. Res. Space Phys.* **101**, 10749–10754, <https://doi.org/10.1029/96JA00323> (1996).
- ⁷⁰S. P. Gary and K. Nishimura, “Resonant electron firehose instability: Particle-in-cell simulations,” *Phys. Plasmas* **10**, 3571–3576 (2003).
- ⁷¹S. P. Gary and H. Karimabadi, “Linear theory of electron temperature anisotropy instabilities: Whistler, mirror, and Weibel,” *J. Geophys. Res. Space Phys.* **111**, A11224, <https://doi.org/10.1029/2006JA011764> (2006).
- ⁷²P. Hellinger, P. Trávníček, J. C. Kasper, and A. J. Lazarus, “Solar wind proton temperature anisotropy: Linear theory and WIND/SWE observations,” *Geophys. Res. Lett.* **33**, L09101, <https://doi.org/10.1029/2006GL025925> (2006).
- ⁷³A. Subel, Y. Guan, A. Chattopadhyay, and P. Hassanzadeh, “Explaining the physics of transfer learning in data-driven turbulence modeling,” *PNAS Nexus* **2**, pgad015 (2023).
- ⁷⁴H. Frezat, J. Le Sommer, R. Fablet, G. Balarac, and R. Lguensat, “A posteriori learning for quasi-geostrophic turbulence parametrization,” *J. Adv. Model. Earth Syst.* **14**, e2022MS003124 (2022).
- ⁷⁵A. A. Kaptanoglu, J. L. Callahan, A. Aravkin, C. J. Hansen, and S. L. Brunton, “Promoting global stability in data-driven models of quadratic nonlinear dynamics,” *Phys. Rev. Fluids* **6**, 094401 (2021).
- ⁷⁶A. J. Christlieb, M. Ding, J. Huang, and N. A. Krupansky, “Hyperbolic machine learning moment closures for the BGK equations,” *Multiscale Modeling Simul.* **23**, 187–217 (2025).
- ⁷⁷J. Pathak, S. Subramanian, P. Harrington, S. Raja, A. Chattopadhyay, M. Mardani, T. Kurth, D. Hall, Z. Li, K. Azizzadenesheli, P. Hassanzadeh, K. Kashinath, and A. Anandkumar, “FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators,” [arXiv:2202.11214](https://arxiv.org/abs/2202.11214) (2022).
- ⁷⁸H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, and J.-N. Thépaut, “The ERA5 global reanalysis,” *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
- ⁷⁹A. Lani, M. S. Yalim, and S. Poedts, “A GPU-enabled finite volume solver for global magnetospheric simulations on unstructured grids,” *Comput. Phys. Commun.* **185**, 2538–2557 (2014).
- ⁸⁰G. Miloshevich, T. Passot, and P. L. Sulem, “Modeling imbalanced collisionless Alfvén wave turbulence with nonlinear diffusion equations,” *Astrophys. J. Lett.* **888**, L7 (2020).
- ⁸¹G. Miloshevich, D. Laveder, T. Passot, and P. L. Sulem, “Inverse cascade and magnetic vortices in kinetic Alfvén-wave turbulence,” *J. Plasma Phys.* **87**, 905870201 (2021).
- ⁸²A. Anandkumar, K. Azizzadenesheli, K. Bhattacharya, N. Kovachki, Z. Li, B. Liu, and A. Stuart, “Neural operator: Graph kernel network for partial differential equations,” in *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations* (2020).
- ⁸³Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, “Fourier neural operator for parametric partial differential equations,” [arXiv:2010.08895](https://arxiv.org/abs/2010.08895) (2021).
- ⁸⁴V. Gopakumar, S. Pamela, L. Zanisi, Z. Li, A. Anandkumar, and M. Team, “Fourier neural operator for plasma modelling,” [arXiv:2302.06542](https://arxiv.org/abs/2302.06542) (2023).
- ⁸⁵T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine, “Enforcing analytic constraints in neural networks emulating physical systems,” *Phys. Rev. Lett.* **126**, 098302 (2021).
- ⁸⁶M. C. McGrae-Menge, J. R. Pierce, F. Fiuza, and E. P. Alves, “Embedding physical symmetries into machine-learned reduced plasma physics models via data augmentation,” [arXiv:2506.14048](https://arxiv.org/abs/2506.14048) (2025).
- ⁸⁷G.-J. Both, S. Choudhury, P. Sens, and R. Kusters, “DeepMoD: Deep learning for model discovery in noisy data,” *J. Comput. Phys.* **428**, 109985 (2021).
- ⁸⁸Y. Yang, Y. Shi, M. Wan, W. H. Matthaeus, and S. Chen, “Energy cascade and its locality in compressible magnetohydrodynamic turbulence,” *Phys. Rev. E* **93**, 061102 (2016).
- ⁸⁹Y. Yang, W. H. Matthaeus, Y. Shi, M. Wan, and S. Chen, “Compressibility effect on coherent structures, energy transfer, and scaling in magnetohydrodynamic turbulence,” *Phys. Fluids* **29**, 035105 (2017).
- ⁹⁰G. Arrò, H. Li, and W. H. Matthaeus, “Spatiotemporal energy cascade in three-dimensional magnetohydrodynamic turbulence,” *Phys. Rev. Lett.* **134**, 235201 (2025).
- ⁹¹W. H. Matthaeus, “Turbulence in space plasmas: Who needs it?,” *Phys. Plasmas* **28**, 032306 (2021).