Advanced surrogate model for electron-scale turbulence in tokamak pedestals

Ionuţ-Gabriel Farcaş [1,2,†], **Gabriele Merlo** [1,3,4] **and Frank Jenko** [1,3,4]

[1]Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, TX 78712, USA

[2]Department of Mathematics, Virginia Tech, VA 24061, USA

[3]Max Planck Institute for Plasma Physics, Boltzmannstr. 2, 85748 Garching, Germany

[4]Institute for Fusion Studies, The University of Texas at Austin, TX 78712, USA

We derive an advanced surrogate model for predicting turbulent transport at the edge of tokamaks driven by electron temperature gradient (ETG) modes. Our derivation is based on a recently developed sensitivity-driven sparse grid interpolation approach for uncertainty quantification and sensitivity analysis at scale, which informs the set of parameters that define the surrogate model as a scaling law. Our model reveals that ETG-driven electron heat flux is influenced by the safety factor q, electron beta β_e and normalized electron Debye length λ_D, in addition to well-established parameters such as the electron temperature and density gradients. To assess the trustworthiness of our model's predictions beyond training, we compute prediction intervals using bootstrapping. The surrogate model's predictive power is tested across a wide range of parameter values, including within-distribution testing parameters (to verify our model) as well as out-of-bounds and out-of-distribution testing (to validate the proposed model). Overall, validation efforts show that our model competes well with, or can even outperform, existing scaling laws in predicting ETG-driven transport.

---

1. Introduction

Turbulent transport is known to determine the energy confinement time of fusion devices. Quantifying, predicting and controlling this turbulent transport is a prerequisite for designing optimized fusion power plants and is therefore considered a key open problem in fusion research. While high-fidelity numerical simulations based on first principles are crucial for understanding the complex mechanisms behind turbulent transport, they are often computationally too expensive for routine use in tasks like optimization or uncertainty quantification that involve large ensembles of simulations.

† Email address for correspondence: ionut.farcas@austin.utexas.edu

Constructing computationally cheap yet reliable surrogate models is therefore highly desirable in practice.

In Farcaş, Merlo & Jenko (2022), we proposed a sensitivity-driven dimension-adaptive sparse grid framework for uncertainty quantification (UQ) and sensitivity analysis (SA) at scale, which was applied to the problem of turbulent transport in the pedestal region of a tokamak driven by electron temperature gradient (ETG) modes. The sensitivity-driven approach intrinsically provides an interpolation-based surrogate model, which turned out to provide accurate predictions for the scenario studied in Farcaş *et al.* (2022) for testing parameters within the interpolation bounds. However, since polynomial extrapolation is, in general, ill posed and unstable (Trefethen 2012), we will not use this model for predictions outside the interpolation bounds used for its construction. In addition, this interpolation model does not incorporate a quantification of prediction uncertainty, which is crucial for assessing the trustworthiness of surrogate transport models for any given input parameters. The present paper builds on Farcaş *et al.* (2022) and leverages the aforementioned sensitivity-driven approach to derive a generic and parsimonious surrogate transport model for ETG-driven turbulent fluxes that (i) depends on physically interpretable parameters, (ii) generalizes beyond training and, importantly, (iii) incorporates a quantification of prediction uncertainty.

Electron temperature gradient turbulence has been the subject of theoretical and numerical investigations for more than two decades, from the seminal works by Jenko *et al.* (2000), Dorland *et al.* (2000) and Jenko & Dorland (2002) focused on core plasmas to more recent studies (Idomura 2006; Told *et al.* 2008; Jenko *et al.* 2009; Hatch *et al.* 2015, 2017; Parisi *et al.* 2020, 2022; Hassan *et al.* 2021; Belli, Candy & Sfiligoi 2022, 2024; Chapman-Oplopoiou *et al.* 2022; Stimmel *et al.* 2022; Leppin *et al.* 2023; Li *et al.* 2023) indicating its role in regulating transport in the pedestal. Given its importance, several recent papers such as Chapman-Oplopoiou *et al.* (2022), Guttenfelder *et al.* (2021) and Hatch *et al.* (2022, 2024) have formulated surrogate models as well as simple algebraic expressions for ETG fluxes in the pedestal.

Here, we propose a scaling law for the ETG-driven electron heat flux as a function of the safety factor q, the electron beta β_e and the normalized electron Debye length λ_D, in addition to well-established parameters such as electron temperature and density gradients. The exponents appearing in our scaling law are obtained by performing a simple regression fit using the existing high-fidelity, nonlinear simulation results from Farcaş *et al.* (2022). To the best of our knowledge, the dependence on q and β_e is new and describes the effect of the magnetic geometry on ETG modes. Furthermore, we incorporate a quantification of uncertainty in the predictions issued by the proposed surrogate model, a critical requirement in data-driven modelling, essential for evaluating the predictive performance for arbitrary input parameters. To this end, we compute prediction intervals via bootstrapping (we refer the reader to Efron & Tibshirani (1986) for a review of bootstrapping methods), which offers a distribution-free and reliable method to account for data variability and model uncertainty. We test the prediction capabilities of the proposed surrogate model across a wide range of parameter values. The model is verified against 32 within-distribution testing parameters and validated against 61 out-of-distribution and 40 out-of-bounds testing parameters. We also compare it with similar scaling laws available in the literature, obtaining similar or more accurate predictions.

The remainder of this paper is organized as follows. Section 2 summarizes the baseline simulation parameters used to derive the proposed surrogate model. Section 3 details the steps used to derive our surrogate model, including how to incorporate a quantification of uncertainty in its predictions by computing prediction intervals via bootstrapping. Section 4 presents our results. We assess the prediction capabilities of the proposed model

and compute prediction intervals via bootstrapping over a wide range of parameter values, from within-distribution testing data to out-of-bounds and out-of-distribution testing parameters. Section 5 investigates the origin of the parametric dependencies appearing in the proposed model, in particular q and β_e. Finally, some overall conclusions are presented in § 6. The code and data to reproduce our results are publicly available at https://github.com/ionutfarcas/surrogate_model_etg_pedestal.

2. Baseline simulation parameters

The present paper builds on the UQ and SA study performed in Farcaş *et al.* (2022) and derives a generic and parsimonious surrogate transport model for ETG turbulence. In the following, we present the aspects relevant for this work and refer the reader to Farcaş *et al.* (2022) for further details.

The scenario under consideration models DIII-D conditions similar to Hassan *et al.* (2021) and Walker & Hatch (2023) and is representative of typical pedestal conditions, where the large gradients can drive a substantial electron heat flux via ETG turbulence; we refer the reader to Hassan *et al.* (2021) for a detailed description of the experimental conditions. We consider that the plasma behaviour is fully specified by the following eight local parameters: $\{n_e, T_e, \omega_{n_e}, \omega_{T_e}, q, \hat{s}, \tau, Z_{\text{eff}}\}$. Here, n_e is the electron density and T_e is the electron temperature, $\omega_{n_e} = a/L_{n_e} = 1/n_e \, dn_e/d\rho_{\text{tor}}$ and $\omega_{T_e} = a/L_{T_e} = 1/T_e \, dT_e/d\rho_{\text{tor}}$ are the respective normalized (with respect to the minor radius a) gradients, where ρ_{tor} denotes the square root of the normalized toroidal magnetic flux, q is the safety factor, $\hat{s} = r/q \, dq/dr$ is the magnetic shear (r denotes the minor radius defined below), $\tau = Z_{\text{eff}} T_e/T_i$ and Z_{eff} is the effective ion charge retained in the collisions operator (here, a linearized Landau–Boltzmann operator). The eight parameters are modelled as independent uniform random variables with symmetric bounds around their nominal values; for more details, see Appendix A. Moreover, the ions are assumed to be adiabatic, and electromagnetic effects, computed consistently with the values of the electron temperature and density, are retained.

The magnetic geometry is specified according to a generalized Miller parametrization (Candy 2009). We employed 64 Fourier harmonics to achieve a sufficiently accurate representation of the flux surface at $\rho_{\text{tor}} = 0.95$, near the pedestal top. Using a generalized Miller parametrization instead of a magnetohydrodynamics (MHD) equilibrium allows us to vary independently (and self-consistently) q and $\hat{s}$. Note that the parametrization of the magnetic geometry is typically affected by uncertainties as well and should therefore be incorporated into the underlying set of uncertain inputs. This, however, would increase the number of uncertain inputs drastically, making the UQ and SA study significantly more challenging; we leave such an extended analysis for future work.

The simulations in Farcaş *et al.* (2022) were carried out using the plasma micro-turbulence simulation code GENE (Jenko *et al.* 2000) in the flux-tube limit using a box with $n_{k_x} \times n_{k_y} \times n_z \times n_{v_\parallel} \times n_\mu = 256 \times 24 \times 168 \times 32 \times 8$ degrees of freedom in the five-dimensional position–velocity space. This grid was shown to be sufficiently fine by performing convergence tests conducted for several input configurations, including the four pairs comprising the gradients extrema, corresponding to the highest and lowest electron heat flow values within the considered parameter range. We note that several works in the literature (Guttenfelder *et al.* 2021; Chapman-Oplopoiou *et al.* 2022; Walker & Hatch 2023) observed that the computed values of ETG fluxes depend sensitively on the resolution n_z in the parallel direction. A resolution comprising $n_z = 168$ points was sufficient in our case because we leveraged the input parameter edge_opt in GENE, which allows use of a lower resolution by redistributing the grid points in the parallel direction. Subsequent simulations in this paper utilize this grid resolution.

3. Parsimonious surrogate model with prediction uncertainty quantification

The sensitivity-driven approach in Farcaş *et al.* (2022) enabled an efficient UQ and SA in the ETG scenario under consideration at a cost of only 57 nonlinear GENE simulations. This low number of simulations was due to the fact that this approach is adaptive, with a refinement indicator based on sensitivity information about the uncertain inputs. In this way, the adaptive procedure preferentially refines the directions corresponding to important input parameters and interactions thereof. Our goal here is to derive an interpretable and predictive surrogate model for ETG turbulent transport with quantified prediction uncertainty. This will be achieved by leveraging information provided by the sensitivity-driven approach, the readily available 57 simulation results and bootstrapping for computing prediction intervals.

A byproduct of the sensitivity-driven approach is that it also intrinsically provides an interpolation-based polynomial surrogate of the output of interest in terms of the uncertain inputs (which can be trivially mapped to a spectral projection basis, a Legendre basis in our case). The sparse interpolation surrogate in Farcaş *et al.* (2022) approximated the power crossing the flux surface in MW due to ETG turbulence in terms of $\{n_e, T_e, \omega_{n_e}, \omega_{T_e}, q, \hat{s}, \tau, Z_{\text{eff}}\}$. As it turns out, this surrogate is accurate for within-distribution testing points, that is, eight-dimensional input parameters falling within the considered uncertainty bounds. For example, the work in Farcaş *et al.* (2022) showed that the mean-squared approximation error at $N = 32$ pseudo-random testing parameters was in $\mathcal{O}(10^{-4})$. The goal of the present paper is to obtain a general surrogate model that is predictive for parameter values beyond the considered uncertainty bounds and, importantly, also incorporates a measure of prediction uncertainty. Since it is well established that polynomial extrapolation is generally ill posed and unstable (Trefethen 2012), we will not exploit the sparse grid surrogate but target a more compact scaling law that relates the turbulent flux to key plasma parameters. The results provided by the sensitivity-driven approach will guide us in choosing which plasma parameters should define the target scaling law.

Prior to describing how the proposed surrogate model is obtained, we address two important details. The initial point of clarification pertains to units and normalizations. A surrogate model in SI units would be preferable because these units are the most general. However, we cannot construct such a model here without redoing all gyrokinetic simulations from Farcaş *et al.* (2022) since the set of uncertain inputs does not include a macroscopic length L_{ref} nor a magnetic field strength B_{ref}. Even if these two parameters would turn out to be unimportant, such an assessment cannot be made *a priori* without redoing the UQ and SA study. We instead opt for a model for the electron heat flux in gyroBohm (GB) units, $Q_{\text{GB}} = n_e T_e^{5/2} m_i^{1/2}/(eB_{\text{ref}} L_{\text{ref}})^2$, which provide a natural normalization for our set-up. This implies that we must map the original 57 simulation results in SI units to their associated heat fluxes in GB units, Q_e/Q_{GB}. Since the flux-surface area is constant, no particular issue arises from removing it from the existing 57 simulation results. However, the GB units depend on two out of the eight uncertain inputs, namely n_e and T_e. Because of this, there is no guarantee that the adaptive procedure in the sensitivity-driven approach would produce the same 57 simulation results in GB units as in the original SI units. Nevertheless, the existing simulation results can be reused irrespective of the units and, as we will show in our results, using GB units does not, in fact, impact the accuracy of the obtained surrogate model.

The second point of clarification concerns the definition of the normalizing quantities. It is crucial to precisely define reference quantities like L_{ref} and B_{ref} to prevent inaccurate model predictions that may not align with experimental measurements, for example. The

same applies to input parameters such as density and temperature gradients. A quantity that impacts the values of these parameters is the selection of the radial coordinate. In the following, we assume that the GB units are defined in terms of the magnetic field on the axis, B_0, and the effective minor radius, $a = \sqrt{\Phi_{\mathrm{LCFS}}/B_0}$, where the toroidal flux value at the last closed flux surface (LCFS), Φ_{LCFS}, represents the length parameter. This implies that the radial coordinate in the definition of the gradients is $\rho_{\mathrm{tor}} = \sqrt{\Phi/\Phi_{\mathrm{LCFS}}}$, where Φ denotes the toroidal flux. It is nevertheless important to keep in mind that these choices directly impact the surrogate model that we will derive next. This implies, for example, that we cannot rule out the possibility that using different definitions, such as alternative radial coordinates, may result in a simpler surrogate (Hatch *et al.* 2024).

Given the choice of normalizing quantities and radial coordinate specified above, we must decide next which input parameters should specify our surrogate model. We seek a set of parameters that are both important, in the sense that variations in these parameters lead to non-negligible variations in the electron heat flux, as well as physically interpretable. To this end, we leverage the information provided by the sensitivity-driven approach. More specifically, we use the fact that this approach provides an accurate sparse grid interpolation surrogate, which, in turn, can be employed to assess the dependence of the heat flux in terms of subsets of uncertain inputs. Using its representation in the Legendre basis, this surrogate, in GB units, amounts to a multi-variate polynomial of third degree comprising 47 terms; its expression is provided in (B1) in Appendix B. Figure 1 plots the one-dimensional dependencies of Q_e/Q_{GB} in terms of all eight uncertain inputs, obtained by fixing the seven other parameters to their respective nominal values. We also perform regression fits that give the rates at which the heat flux varies with each parameter. We remark that these dependencies do not account for parameter interactions (such interactions would appear in higher-dimensional dependencies). Moreover, some rates may change if the remaining seven parameters (e.g. the two gradients) were fixed to other values, such as their extrema.

Nonetheless, the obtained results indicate that the two gradients lead to the most significant dependencies, which is in line with what was expected. In contrast, the dependencies in terms of $\hat{s}$ and Z_{eff} are clearly negligible, implying that $\hat{s}$ and Z_{eff} can be ignored in our surrogate model. We observe small but non-negligible dependencies due to $\{\tau, q, n_e, T_e\}$. Thus, these results suggest a dependence on $\{\omega_{T_e}, \omega_{n_e}, q, \tau, n_e, T_e\}$. Naively using these input parameters is not desirable since n_e and T_e are not easily interpretable. Instead, by examining the gyrokinetic equations, we identify interpretable parameters depending on n_e or T_e that may enter our model. These are the collisionality (measured by an appropriate collision frequency ν_c), Debye screening (described by the normalized electron Debye length λ_D; in the following we will use $\lambda_D = \lambda_{De}/\rho_e$ normalized with the electron Larmor radius ρ_{e}, the natural microscopic length scale for our problem) and plasma beta $\beta_e = 403 \times 10^{-5} n_e T_e/B_0$ (here, the electron density n_e is measured in $10^{19}/m^3$, T_e in keV and B_0 in T), which affects the magnetic geometry and causes magnetic fluctuations. However, since we cannot perform a simple change of variables from $\{n_e, T_e\}$ to $\{\lambda_D, \nu_c, \beta_e\}$, we instead perform additional simulations to identify which subset of $\{\lambda_D, \nu_c, \beta_e\}$ should enter our surrogate model. We note that considering $\{\lambda_D, \nu_c, \beta_e\}$ instead of $\{n_e, T_e\}$ allows us to more clearly pinpoint possible physical effects, and leads to a more interpretable and intuitive surrogate model.

We perform additional scans in which collisions, Debye shielding or electromagnetic effects are individually switched off. For a comprehensive perspective, we perform these scans for both left and right uniform bounds of $\{\omega_{T_e}, \omega_{n_e}, q, \tau, n_e, T_e\}$ used in the sensitivity-driven approach (corresponding to their respective smallest and largest values; see table 1 in Appendix A). Figure 2 plots the results. We observe a clear dependence on
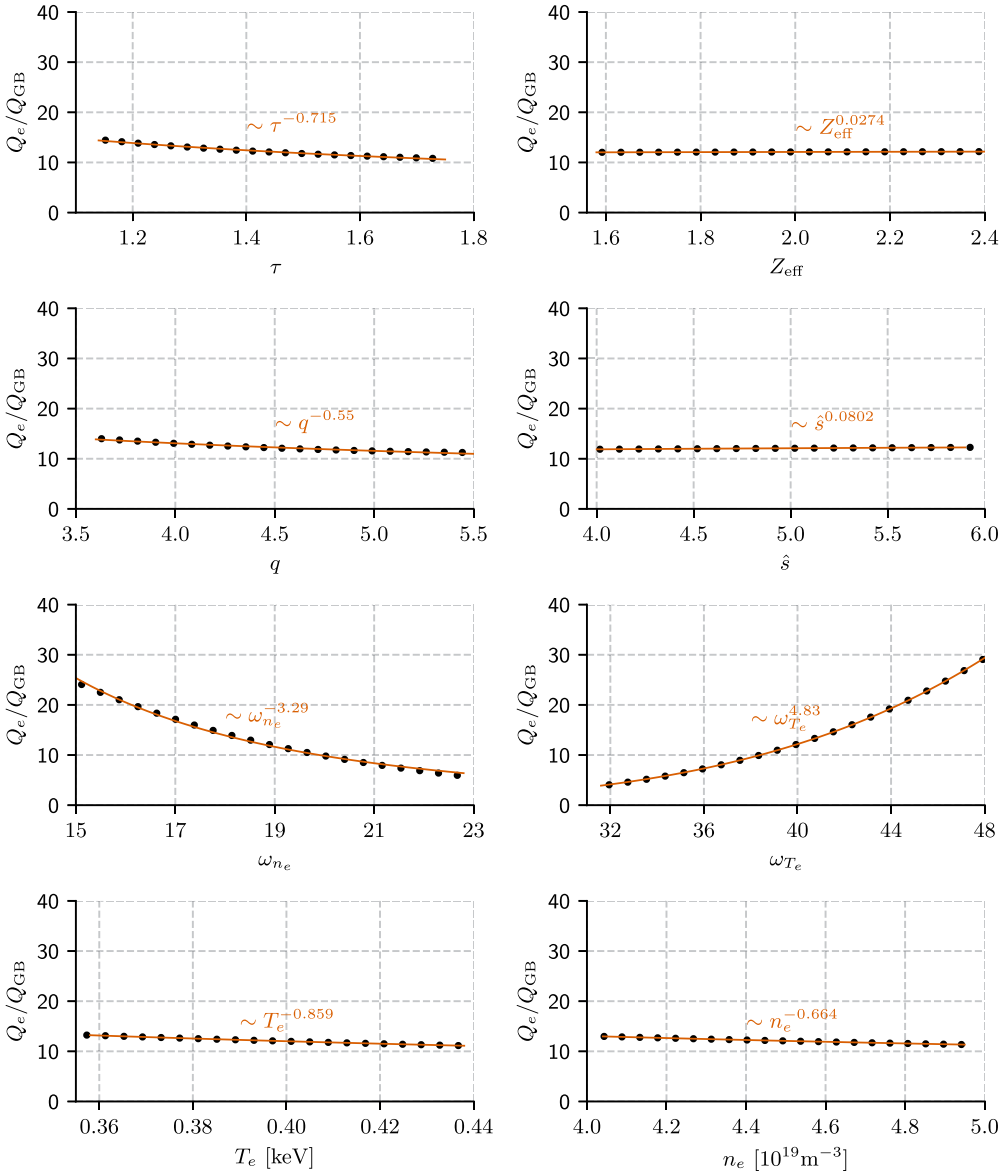
FIGURE 1. Dependence of the electron heat flux on each of the eight considered parameters obtained using the sparse grid surrogate model. The remaining seven parameters are fixed to their respective nominal values. We also estimate via regression the rates at which the flux varies with the eight inputs.

β_e and a weaker but not negligible one on λ_D. Collisions, in contrast, do not lead to any significant variations in the heat fluxes, which implies that ν_c is unimportant. Based on these results, our surrogate model should depend on $\{\omega_{T_e}, \omega_{n_e}, q, \tau, \beta_e, \lambda_D\}$. Lastly, since it is well established that ETG transport is a threshold process with respect to $\eta_e = \omega_{T_e}/\omega_{n_e}$, with finite fluxes only when $\eta_e \gtrsim 1$ as shown in Jenko, Dorland & Hammett (2001) and Jenko *et al.* (2000), we use η_e instead of ω_{n_e}. We therefore conclude that the input parameters that define our surrogate model are $\{\omega_{T_e}, \eta_e, q, \tau, \beta_e, \lambda_D\}$.
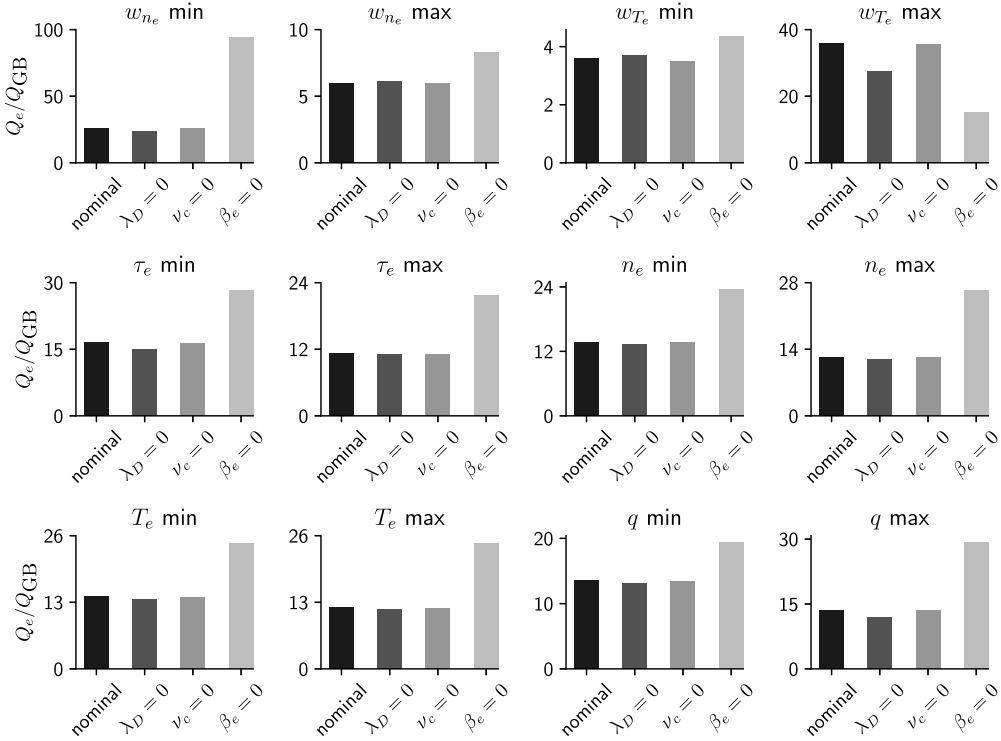
FIGURE 2. Dependence of the electron heat flux on various physical effects. Each panel compares the nominal heat flux with the one obtained when, individually, Debye shielding (λ_D), collisions (ν_c) or electromagnetic effects (β_e) are excluded. In each case, all other plasma parameters except the one indicated in the title are kept to their respective nominal values.

We seek a surrogate model as a scaling law of the form

$$Q_e/Q_{GB} = c_0 \sqrt{\frac{m_e}{m_i}} \omega_{T_e}^{p_1} (\eta_e - 1)^{p_2} \tau_e^{p_3} q^{p_4} \beta_e^{p_5} \lambda_D^{p_6},\tag{3.1}$$

where $\{c_0, p_1, p_2, p_3, p_4, p_5, p_6\} \in \mathbb{R}^7$. To determine these coefficients, we perform a data fit using the existing 57 nonlinear GENE simulations from our UQ and SA study from Farcaş *et al.* (2022), where we map the original inputs $\{\omega_{T_e}, \omega_{n_e}, q, \tau, n_e, T_e\}$ to $\{\omega_{T_e}, \eta_e, q, \tau, \beta_e, \lambda_D\}$ and the corresponding outputs in SI units to heat fluxes in GB units. We note that fitting the power law (3.1) directly is not straightforward as the target function is nonlinear. Even more, a poor choice of the loss function or minimization approach can lead to a biased model that overfits and hence generalizes poorly. We take advantage of the fact that we want to fit a scaling law with strictly positive parameters and perform a standard regression fit in logarithmic coordinates. We obtain

$$Q_e/Q_{GB} = 0.038 \sqrt{\frac{m_e}{m_i}} \omega_{T_e}^{1.40} (\eta_e - 1)^{1.79} \tau_e^{-0.76} q^{-0.51} \beta_e^{-0.87} \lambda_D^{-0.51}.\tag{3.2}$$

To assess the validity of the predefined threshold $\eta_{e,0} = 1$ in the proposed model, we perform an additional experiment in which $\eta_{e,0}$ is a free parameter. That is, we consider

the surrogate model

$$Q_e/Q_{\mathrm{GB}} = c_0' \sqrt{\frac{m_e}{m_i}} \omega_{T_e}^{p_1'} (\eta_e - \eta_{e,0})^{p_2'} \tau_e^{p_3'} q^{p_4'} \beta_e^{p_5'} \lambda_D^{p_6'}, \tag{3.3}$$

where $\{c_0', p_1', \eta_{e,0}, p_2', p_3', p_4', p_5', p_6'\} \in \mathbb{R}^8$. We determine these coefficients analogously to (3.1) and obtain

$$Q_e/Q_{\mathrm{GB}} = 0.048 \sqrt{\frac{m_e}{m_i}} \omega_{T_e}^{1.40} (\eta_e - 1.03)^{1.70} \tau_e^{-0.76} q^{-0.52} \beta_e^{-0.85} \lambda_D^{-0.41}. \tag{3.4}$$

Model (3.4) is similar to the model in (3.2) and in fact both models yield almost identical predictions for all datasets considered in § 4. This fit therefore reveals that the threshold $\eta_{e,0} = 1$ is valid in our context. We note, however, that we cannot disregard the fact that $\eta_{e,0}$ may have a different value. Assessing this would require performing simulations with η_e values close to the threshold, which, in our context, would again necessitate redoing the sparse grid study since the minimum η_e value for the 57 sparse grid points was $\eta_{e,\mathrm{min}} = 1.4$. We leave such an analysis for future research.

In practice, the deterministic predictions from the model in (3.2) are not enough. Generally, in order to enhance the reliability of surrogate models, particularly those derived from data fitting, it is important for their predictions to include an element of prediction uncertainty. To tackle this issue, we calculate prediction intervals using bootstrapping. Bootstrapping offers a reliable method for establishing prediction intervals without relying on restrictive assumptions regarding the distribution of data and errors. To compute prediction intervals, we resample the 57 input–output pairs $M \in \mathbb{N}$ times with replacement and fit a model for each of the M resampled pairs via regression, as described above. We then compute the target predictions for each fitted model, which results in an ensemble of M predictions. Lastly, this ensemble is used to calculate prediction intervals.

4. Predictions using the novel surrogate model

We now assess the prediction capabilities of the novel surrogate model given in (3.2) and compute prediction intervals using bootstrapping. For a comprehensive perspective, we perform three sets of experiments comprising 133 testing points in total across a wide range of values, including two sets that go beyond the training data.

We will compare our model with the most accurate surrogate proposed in Hatch *et al.* (2022), obtained via symbolic regression using a database of $N = 61$ ETG simulations comprising discharges from the JET, DIII-D, ASDEX Upgrade and C-MOD tokamaks

$$Q_e/Q_{\mathrm{GB}} = \sqrt{\frac{m_e}{m_i}} \omega_{T_e} (1.44 + 0.50 \eta_e^4). \tag{4.1}$$

To the best of our knowledge, this represents the most comprehensive such database currently available in the literature. The more recent work in Hatch *et al.* (2024) refined the formulas from Hatch *et al.* (2022) and proposed new, richer models that provide more accurate predictions for the database. We therefore additionally consider the model given in (1) in Hatch *et al.* (2024)

$$Q_e'/Q_{\mathrm{GB}} = 0.019 \sqrt{\frac{m_e}{m_i}} (\omega_{T_e}')^2 (\eta_e - 1) \eta_e^{1.57} \tau_e^{-0.5} (\lambda_D')^{-0.4}, \tag{4.2}$$

where the quantities marked with a prime symbol are defined differently than in the present paper. More specifically, ω_{T_e}' denotes the temperature gradient taken with respect

to the normalized poloidal flux, ψ, and λ'_D denotes the Debye length normalized to $\rho_s = \sqrt{T_e/m_i}m_i/eB$. In addition, the evaluation of the heat flux uses the effective flux-surface area $A = 2\pi R 2\pi a$, where R denotes the major radius and a the minor radius (for our case, $R = 1.68$ m and $a = 0.77$ m) instead of the true flux surface area V'. As noted in § 3, the conversion factors between radial coordinates has some dependence on q and β_e. We therefore cannot exclude the possibility that a model formulated for the radial coordinate r (as in this paper) has different q and β_e dependence than a model formulated for the radial coordinate ψ as in (4.2).

To use the model in (4.2) in our context, we must (i) convert our temperature gradient and Debye length to be compatible with the units used in (4.2) and (ii) convert the ensuing heat flux predictions to be compatible with our definition by accounting for the different flux-surface area. The conversion of the Debye length is straightforward, namely $\lambda'_D = \lambda_D \sqrt{m_e/m_i}$. The conversion for the temperature gradient and heat flux is explained in detail in Appendix D. As noted there, a full conversion of the temperature gradient is not possible because it requires knowledge of a global MHD equilibrium. Nonetheless, we employ the model in (4.2) in our experiments both for a more comprehensive perspective and also because this model is richer than the model in (4.1). For more details about model (4.2), we refer the reader to Hatch *et al.* (2024). We furthermore note that other works such as Chapman-Oplopoiou *et al.* (2022) and Guttenfelder *et al.* (2021) provide algebraic trends of ETG fluxes, which, however, do not represent general surrogate ETG models as their expressions depend on undetermined constants. The work in Hatch *et al.* (2024) showed that the models proposed therein, including (4.2), yield more accurate predictions than the aforementioned algebraic expressions with constants fitted using the database in Hatch *et al.* (2022).

In all our experiments, we use $M = 1000$ bootstrapping samples and compute 95% prediction intervals. Furthermore, to assess the prediction accuracy of the deterministic predictions issued by our model (i.e. without taking into account the prediction intervals) in a manner consistent with common practice in the literature, we will utilize the following error metric that compares a set of N reference electron heat fluxes (hereby denoted by $Q_{e,\text{ref}}/Q_{\text{GB}}$) and corresponding (deterministic) approximations obtained via a surrogate model (denoted by $Q_{e,\text{approx}}/Q_{\text{GB}}$), $\{Q_{e,\text{ref};i}/Q_{\text{GB}}, Q_{e,\text{approx};i}/Q_{\text{GB}}\}_{i=1}^{N}$:

$$\varepsilon(Q_{e,\text{ref}}/Q_{\text{GB}}, Q_{e,\text{approx}}/Q_{\text{GB}}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \frac{(Q_{e,\text{ref};i}/Q_{\text{GB}} - Q_{e,\text{approx};i}/Q_{\text{GB}})^2}{(Q_{e,\text{ref};i}/Q_{\text{GB}} + Q_{e,\text{approx};i}/Q_{\text{GB}})^2}}. \tag{4.3}$$

This error metric was also adopted in Hatch *et al.* (2022, 2024) and it equally penalizes extreme cases, that is, $Q_{e,\text{approx}} \ll Q_{e,\text{ref}}$ and $Q_{e,\text{approx}} \gg Q_{e,\text{ref}}$.

4.1. *Within-distribution testing parameters*

We initially investigate whether the proposed regression-based scaling law (3.2) exhibits any significant decrease in accuracy compared with the sensitivity-driven sparse grid surrogate model obtained from Farcaş *et al.* (2022), which was shown to be accurate for within-distribution testing data. For this purpose, we utilize the $N = 32$ testing samples from Farcaş *et al.* (2022). We map the original input parameters $\{\omega_{T_e}, \omega_{n_e}, q, \tau, n_e, T_e\}$ to $\{\omega_{T_e}, \eta_e, q, \tau, \beta_e, \lambda_D\}$ and the corresponding outputs to heat fluxes in GB units. In addition, we also assess the predictions provided by models (4.1) and (4.2).

Figure 3 shows the results. To simplify visualization, we reordered the results in ascending order relative to the reference results. Panel (*a*) shows the predictions obtained using our surrogate model and the sparse grid model based on the work in Farcaş *et al.*
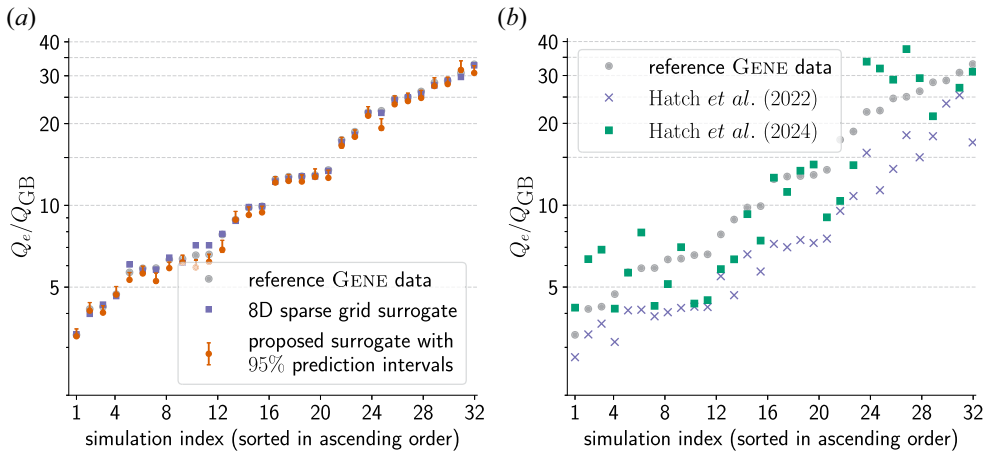
FIGURE 3. (*a*) Comparison between the proposed surrogate model (with 95 % prediction intervals) and the more complex sparse grid surrogate model depending on all eight uncertain inputs from Farca\u015f *et al.* (2022) in GB units using $N = 32$ within-distribution testing data points. (*b*) The corresponding predictions using models (4.1) and (4.2). To simplify visualization, we reordered the fluxes in ascending order relative to the reference values.

(2022). Both surrogate models provide accurate predictions that are close to each other. In addition, the 95 % prediction intervals, shown for each heat flux predicted by our model, are small. The corresponding deterministic errors are $\varepsilon = 0.0145$ for the sparse grid model and $\varepsilon = 0.0292$ for the proposed surrogate model. We can therefore conclude that the proposed regression-based scaling law (3.2) provides accurate predictions with small prediction intervals that do not deteriorate the accuracy of the more complex sparse grid surrogate for within-distribution testing data. Panel (*b*) shows the predictions obtained with models (4.1) and (4.2). The model in (4.1) slightly underpredicts the reference fluxes, but nevertheless yields an error $\varepsilon = 0.2248$. The richer model (4.2) improves these predictions, decreasing the error to $\varepsilon = 0.1561$.

4.2. *Testing the model using data beyond training*

Next, we evaluate the predictive performance of the proposed model on data that fall outside the input distribution. We examine two datasets with parameters that exceed the training boundaries. The first dataset consists of 61 out-of-distribution testing points from the database in Hatch *et al.* (2022), while the second dataset includes 40 out-of-bounds testing points. These experiments serve as validation tests for our model.

4.2.1. *Out-of-distribution testing parameters*

We test the proposed model using the database of $N = 61$ ETG simulations from Hatch *et al.* (2022). Given that the 61 entries pertain to four distinct machines and different configurations, and none of the input pairs $\{\omega_{T_e}, \omega_{n_e}, \tau, \beta_e, \lambda_D\}$ from the database fall within the uniform bounds of our inputs, it is reasonable to consider these 61 data points as out-of-distribution.

Figure 4 plots the results. This experiment highlights the benefits of including prediction intervals: these intervals not only indicate the level of uncertainty in our predictions but also enhance our understanding of the predictive capabilities of our model. Specifically, 17 reference fluxes are within the 95 % prediction intervals of our model forecasts, and most of the other reference fluxes are close to our prediction intervals, with the exception of a handful of outliers. There are five outliers for which the respective relative error exceeds
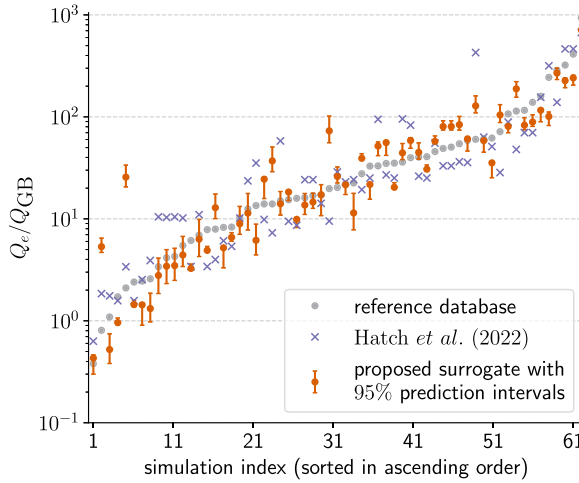
FIGURE 4. Comparison between the proposed surrogate model (with 95 % prediction intervals) and the surrogate model (4.1) at the $N = 61$ data points from the database in Hatch *et al.* (2022). These points represent out-of-distribution testing data for our model. The refined surrogate model (4.2) provides more accurate predictions with error $\varepsilon = 0.15$. To simplify visualization, we reordered the fluxes in ascending order relative to the reference values.

100 % (corresponding to indices 16, 17, 19, 20 and 46 in the database in Hatch *et al.* 2022). Correspondingly, the model in (4.1) yields 10 such outliers. The value of the deterministic error (4.3) for our model is $\varepsilon = 0.2604$, which is slightly smaller than the error of the surrogate model given by (4.1), $\varepsilon = 0.2860$, which was trained using the entire database. We note, however, that model (4.2) (the corresponding results are not plotted in figure 4 because the necessary conversion factors are not available in the database in Hatch *et al.* 2022) provides more accurate predictions (in the units used in Hatch *et al.* 2024), reducing the error by nearly 50 % to $\varepsilon = 0.15$.

4.2.2. *Out-of-bounds testing parameters*

Lastly, for a more comprehensive assessment of our proposed surrogate model, we test its prediction accuracy using out-of-bounds testing data. For this experiment, we generate $N = 40$ uniform testing samples that fall outside of the training bounds given in table 1 in Appendix A. The values of the $N = 40$ input parameters as well as the corresponding heat fluxes in GB units are listed in table 2 in Appendix C.

Figure 5 plots the results. Our surrogate model closely matches the reference results. In fact, 35 out of the 40 reference flux values are within or very close to our prediction intervals. The corresponding deterministic error amounts to $\varepsilon = 0.2074$. In contrast, the surrogate given in (4.1) produces fairly inaccurate predictions that overestimate the reference heat fluxes, resulting in a large error $\varepsilon = 0.5451$. We attribute these results to the fact that the model (4.1) is not rich enough for these testing points as it depends on only two parameters, ω_{T_e} and η_e. In contrast, the refined surrogate model (4.2) produces more accurate predictions that decrease the error down to $\varepsilon = 0.3280$.

5. On the scaling parameter dependencies

We examine in more detail the dependence of the new scaling law on the parameters $\{\tau, \lambda_D, \beta_e, q\}$. With this aim, we conduct supplemental simulations in which each of the four parameters is varied individually, with all others held constant at their nominal values.
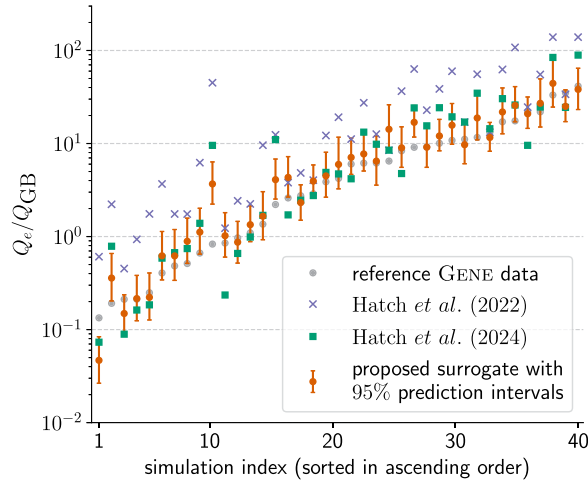
FIGURE 5. Comparison between the predictions obtained using the proposed surrogate model (plus their corresponding 95 % prediction intervals) and surrogates (4.1) and (4.2) at $N = 40$ out-of-bounds testing points. To simplify visualization, we reordered the fluxes in ascending order relative to the reference values.

Figure 6 shows the dependence of the electron heat flux on τ. We explore a substantially broader spectrum of values for this parameter than those used for the construction of the surrogate model, while ensuring these values are realistic and accounting for a possible large impurity content. Our results indicate a reduction in the heat flux, which is consistent with previous findings. For instance, the study by Jenko *et al.* (2001) found that the linear dynamics of ETG modes shares similarities with ion temperature gradient modes, albeit with the electrons and ions being interchanged. Consequently, an increase in the electron-to-ion temperature ratio, τ, is likely to exert a stabilizing influence on ETG-driven fluxes. We compare the reference GENE data with the predictions obtained using our surrogate model in (3.2) (plus their corresponding 95 % prediction intervals) by fixing all parameters except for τ to their respective nominal values in our surrogate. Our predictions in figure 6 not only reflect this relationship but do so with high accuracy. In fact, all reference GENE values fall within the 95 % prediction intervals.

We proceed with examining the impact of varying the normalized electron Debye length λ_D. The corresponding results are plotted in figure 7. Panel (*a*) compares the reference GENE data with the predictions obtained via our surrogate model (plus their corresponding 95 % prediction intervals) for $\lambda_D > 0$ by setting all parameters except for λ_D to their respective nominal values. As λ_D exceeds unity, we observe a decrease in the electron heat flux, corroborating our expectations. For instance, this is consistent with the observed stabilization of the high-k_y modes with increasing Debye length, as illustrated in panel (*b*) of figure 7. The nominal value $\lambda_D = 0.93$ lies within the range where Debye shielding starts to exert a significant influence on transport. We remark that the deviation between GENE and the surrogate predictions for large values of λ_D was expected since the surrogate model was constructed without exploring this region. This, however, does not limit the scope of our model since typical plasma parameters of current and future pedestals do not fall within this region.

Investigating the effects of the remaining two parameters, β_e and q, presents a more considerable challenge due to their multifaceted role in the gyrokinetic equations. Specifically, β_e influences not only the field equations but also the particle dynamics
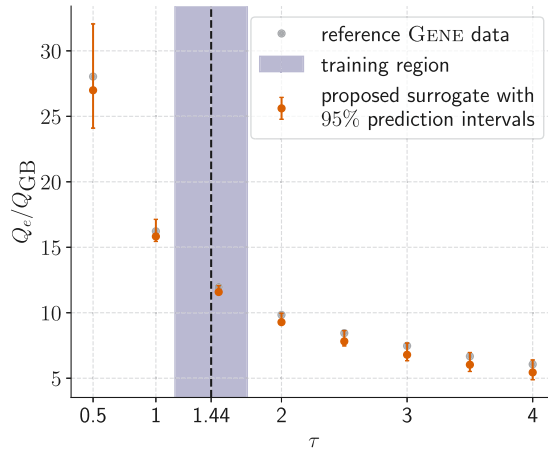
FIGURE 6. Dependence of the electron heat flux on τ values that exceed the training bounds. We compare the reference GENE data with the predictions obtained using our surrogate model (plus their corresponding 95 % prediction intervals).
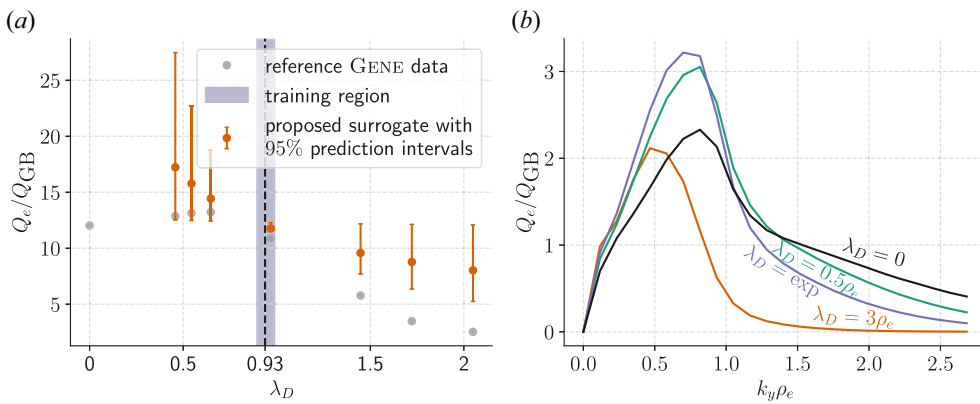

FIGURE 7. (*a*) Dependence of the electron heat flux on $\lambda_D \geq 0$ values that exceed the training bounds. The reference GENE data are compared with the predictions obtained via our surrogate (plus their 95 % prediction intervals) for $\lambda_D > 0$. (*b*) Flux spectra for different values of λ_D.

through the pressure gradient ∇p, which contributes to the drift velocity. Additionally, within the context of magnetic geometry, β_e plays a critical in determining the Shafranov shift, again via the pressure gradient. When working with normalized parameters, we obtain

$$\nabla p = -\beta_e \sum_j \frac{n_j}{n_e} \frac{T_j}{T_e} (\omega_{T,j} + \omega_{n,j}), \qquad (5.1)$$

where the sum runs over all plasma species. Consequently, we conduct parameter scans in which we vary β_e and simultaneously compute all the affected terms. We then compare the results with those from scenarios in which we maintain a constant pressure gradient to isolate and evaluate the impact of the magnetic geometry on its own, as well as in conjunction with the curvature drift effects. The results of this comprehensive analysis are presented in figure 8. This reveals that the behaviour of the turbulent fluxes differs depending on the role played by β_e. With a consistently varying β_e (the black line
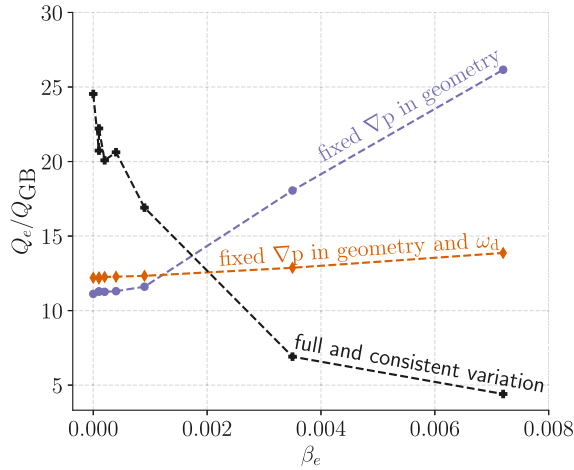
FIGURE 8. Dependence of the electron heat flux on β_e. The black line plots the simulation results where all terms in the gyrokinetic equations are modified consistently. The orange line shows the results in which the pressure gradient ∇p is held constant when evaluating particle drifts and the magnetic equilibrium. The purple line plots the simulation results where ∇p is only kept constant when determining the magnetic equilibrium.

in figure 8), there is a noticeable stabilizing impact on the fluxes. However, when the pressure gradient is held constant (orange line), the influence of β_e appears to be minimal. Additionally, we note an escalation in transport (purple line) when drift velocities are proportionally increased with β_e, indicating that ETG modes are not significantly impacted by dynamical changes stemming from electromagnetic fluctuations. The β_e scaling in our surrogate model, as described by (3.2), should therefore primarily be interpreted as reflective of a pressure gradient scaling. This scaling corresponds to the effect of the Shafranov shift, for which β_e serves as an effective substitute at fixed logarithmic gradients. This interpretation is further corroborated by our finding that the electron heat flux is predominantly electrostatic. The electromagnetic contribution to the heat flux remains consistently minor and exhibits no discernible variation with changes in β_e across the various scenarios we investigated.

In our final analysis, we examine the impact of varying the safety factor q beyond the range considered for training. It is important to note that, while q does not directly appear in the gyrokinetic equations, it is integral to the characterization of the magnetic geometry and therefore indirectly affects all related metric elements. The most significant influence arises from varying the magnetic curvature, as shown in figure 9. Upon increasing the value of q, we observe a reduction of the turbulent fluxes, which eventually plateaus for $q \geq 6$. To better understand this pattern, we compare these results with instances where we purposefully set both the radial (K_x) and binormal (K_y) curvatures to zero. This particular case (orange line in figure 9) reveals a flux dependence similar to that in simulations preserving the complete geometry, although the levels of transport are marginally lower. Such a correlation upholds our hypothesis that the ETG modes in question tend to exhibit slab-like characteristics. The influence of curvature on our system's behaviour is evident in the results where only K_y is considered (purple line in figure 9). Here, the fluxes decrease more rapidly with q, indicating that a higher safety factor causes a stabilizing modification of K_y. This is confirmed in figure 10: for our specific case, the binormal curvature turns positive at all parallel positions once $q > 4$, balancing the destabilizing
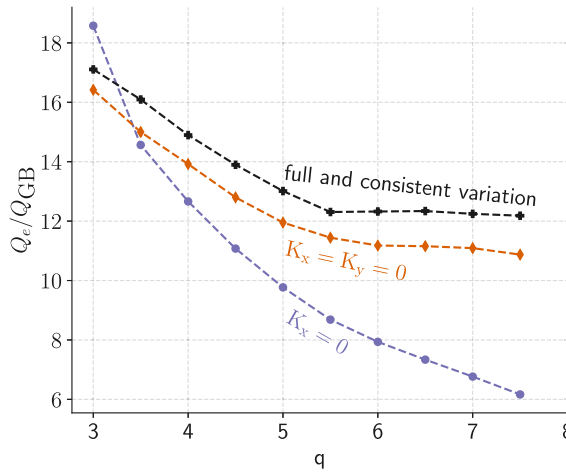
FIGURE 9. Dependence of electron heat flux on q. The black line plots the results achieved when all geometric elements are calculated consistently. The orange line plots the results obtained when both K_x and K_y are artificially set to zero, and the purple line plots the results when only K_x is set to zero.
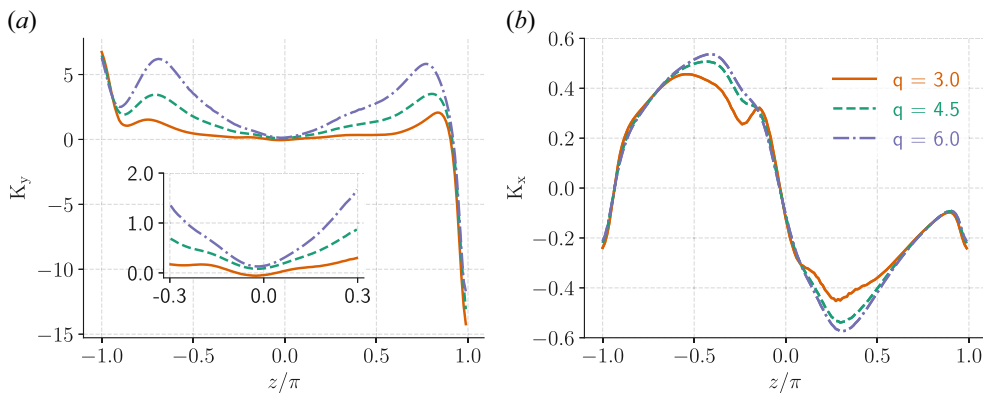

FIGURE 10. Binormal K_y (*a*) and radial K_x (*b*) components of the magnetic curvature as a function of the parallel coordinate z for different values of the safety factor q. The inset in the left plot provides a detailed view of the range $-0.3 < z/\pi < 0.3$, demonstrating that, at sufficiently large values of q, the curvature becomes strictly positive.

influence typically associated with K_x. Throughout all examined scenarios, including those where curvatures were artificially nullified, turbulence remains predominantly clustered around the outboard mid-plane. This consistency points to the prevalent role of finite Larmor radius effects in determining the parallel structure of the ETG modes, irrespective of curvature considerations.

6. Conclusions

In the present work, we employed our recently developed sensitivity-driven dimension-adaptive sparse grid approximation method in the context of ETG turbulence in a tokamak pedestal, characterized by eight uncertain input parameters. This technique enabled us to determine an advanced surrogate model, which takes the form of the

following scaling law:

$$Q_e/Q_{GB} = 0.038\sqrt{\frac{m_e}{m_i}}\omega_{T_e}^{1.40}(\eta_e - 1)^{1.79}\tau_e^{-0.76}q^{-0.51}\beta_e^{-0.87}\lambda_D^{-0.51}, \qquad (6.1)$$

where $Q_{GB} = n_e T_e^{5/2} m_i^{1/2}/(eB_0a)^2$. To our knowledge, the identified dependencies on the safety factor q and plasma beta β_e are novel contributions, elucidating the influence of magnetic geometry on ETG modes. We further enhanced the robustness of our surrogate model by integrating a quantification of uncertainty in its predictions, which is of paramount importance in surrogate modelling. This was achieved through the computation of prediction intervals using the bootstrapping technique. Such measures are essential for evaluating the out-of-sample predictive accuracy of these models, extending their reliability beyond the scope of the training data. We underpinned our approach with extensive numerical evidence, utilizing a total of 133 test points – including data points beyond the training set – to demonstrate that our surrogate model delivers predictions with an acceptable level of precision. The inclusion of prediction intervals not only adds rigour to our model's predictive capabilities but also affords a more profound understanding of its performance. Within the wider framework of turbulent transport simulations in fusion devices, our work implies that sensitivity-driven sparse grid approximations can be effectively harnessed to construct surrogate transport models directly from nonlinear, high-fidelity simulations, presenting a significant advancement in the field.

One application of the proposed surrogate model is profile reconstruction, which can be done similarly to Hatch *et al.* (2024). Starting from an imposed boundary condition at the edge and an initial guess, the surrogate model can be used to provide an approximation of the heat flux that can be then fed to a transport solver evolving the electron profile iteratively until convergence. Such an application is beyond our scope and is therefore left for future work. In addition, further refinement of the surrogate model considering, for example, the value of the critical gradients as well as the impact of toroidal ETG is also left for our future research.

Acknowledgements

Declaration of interests

The authors report no conflict of interest.

Funding

Data availability statement

The data that support the findings of this study are openly available in the repository at https://github.com/ionutfarcas/surrogate_model_etg_pedestal.

Authors contributions

I.-G.F. and G.M. derived the theory, performed the simulations and analysed the data. All authors contributed equally to reaching conclusions and in writing the paper.

Appendix A. Set-up for the eight input plasma parameters

The set-up for the eight parameters describing the base ETG scenario is listed in table 1. This includes their respective nominal values (second column) and the corresponding left and right uniform bounds considered in the UQ and SA analysis in Farcaş *et al.* (2022). We note that, in Farcaş *et al.* (2022), the temperature and density gradients were normalized with respect to the major radius, R, and the output of interest was the electron heat flow computed in SI units (MW).

Appendix B. Eight-dimensional sparse grid surrogate model

The sparse grid polynomial surrogate in GB units with $Q_{\mathrm{GB}} = n_e T_e^{5/2} m_i^{1/2}/(eB_0 a)^2$, depending on all eight uncertain inputs listed in table 1, obtained from our UQ and SA study in Farcaş *et al.* (2022), expressed using a Legendre basis, reads

$$
\begin{aligned}
Q_e/Q_{\mathrm{GB}} = {} & 0.19866664868934378\tau\omega_{n_e}\omega_{T_e} - 1.1413477734069242\tau^2\omega_{n_e} \\
& + 1.1300923036977033\tau^2\omega_{T_e} + 0.012387562637082726\omega_{n_e}^2\omega_{T_e} \\
& - 0.08487445688736855\tau\omega_{T_e}^2 + 0.009721433191939952q\omega_{T_e}^2 \\
& - 0.005966686182384836\omega_{n_e}\omega_{T_e}^2 + 0.0014173787745147169\omega_{T_e}^3 \\
& - 0.010912805921383906\omega_{n_e}^3 - 0.14098572222663125\tau\omega_{n_e}^2 \\
& + 0.0878050918300935q^2\omega_{T_e} - 0.4779566243839404\omega_{n_e}^2 T_e \\
& + 0.40612708864138414\omega_{n_e}\omega_{T_e}T_e - 0.10152212813990472\omega_{T_e}^2 T_e \\
& - 0.09610059439413138q^3 - 6.565071206471347\tau^3 \\
& - 0.007095934720313662\omega_{T_e}^2 n_e + 0.10660808146140052q\omega_{n_e} \\
& + 0.7209441445034814\omega_{n_e}^2 + 2.3708835567864472\tau\omega_{n_e} \\
& - 1.5507824853701202q^2 - 0.6219694465716796\tau q \\
& - 1.5881501546239818\tau\omega_{T_e} + 11.292418391099027\tau^2 \\
& - 1.6038679243023q\omega_{T_e} + 0.37186341930335004n_e^2 \\
& - 0.6593622299176898\omega_{n_e}\omega_{T_e} + 0.16335509513250532\omega_{T_e}^2 \\
& + 0.369546253431427\omega_{T_e}n_e + 5.968742046874542\tau T_e \\
& + 3.3170053895934304qT_e + 7.075768404578394\omega_{n_e}T_e \\
& - 2.5422978283907813\omega_{T_e}T_e + 64.48667130778047T_e^2 \\
& + 1.3682599362541525T_e n_e + 0.2294389248146788\omega_{n_e}n_e \\
& - 0.06963000365198978\tau n_e + 0.3511888827695439qn_e \\
& + 31.28952697970901q - 10.12874179297306\tau \\
& + 0.16860715589513Z_{\mathrm{eff}} - 9.035737970505384\omega_{n_e} \\
& - 14.949102090605091n_e + 0.19518454179911435\hat{s} \\
& - 113.39196370792155T_e + 7.000425385025428\omega_{T_e} \\
& - 11.083912068866233. \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (B1)
\end{aligned}
$$

Uncertain input parameter	Nominal value	Left uniform bound	Right uniform bound
Electron temperature T_e [keV]	0.3970	0.3573	0.4367
Electron density n_e [10^{19}m^{-3}]	4.4923	4.0428	4.9412
Temperature gradient $\omega_{T_e} = a/L_{T_e}$	39.9347	31.9477	47.9216
Density gradient $\omega_{n_e} = a/L_{n_e}$	18.3928	15.1150	22.6726
Temperature ratio τ	1.4400	1.1520	1.7280
Effective ion charge Z_{eff}	1.9900	1.5920	2.3880
Safety factor q	4.5362	3.6289	5.4434
Magnetic shear $\hat{s}$	5.0212	4.0169	6.0254

TABLE 1. Summary of the eight uniform uncertain parameters considered in Farcaş *et al.* (2022). The second column shows their nominal (mean) value. The corresponding left and right uniform bounds are listed, respectively, in the third and fourth columns. The density and temperature gradients are normalized with respect to the minor radius, a.

It amounts to a multi-variate polynomial of third degree, comprising 47 terms.

Appendix C. Out-of-bounds testing data

Table 2 shows the values of the GENE parameters $\{\omega_{T_e}, \eta_e, q, \tau, \beta_e, \lambda_D, \hat{s}, Z_{\text{eff}}\}$ (columns two to nine) as well as the corresponding electron heat fluxes in GB units (last column) for the 40 out-of-bounds testing parameters used in § 4.2.2.

Index	ω_{T_e}	η_e	τ	q	β_e	$\lambda_D = \lambda_{De}/\rho_e$	$\hat{s}$	Z_{eff}	Q_e/Q_{GB}
1	33.0	3.0	0.55	4.0	0.0124	1.25707748	2.0	1.65	9.123
2	31.0	2.0	0.775	3.5	0.01128	1.10252989	4.0	0.975	2.731
3	26.0	4.0	0.325	4.5	0.01128	1.50249497	0.2	2.325	11.56
4	10.62	2.5	0.6625	3.25	0.02606	1.04394603	1.0	1.988	0.4035
5	30.62	3.5	0.8875	4.75	0.00784	1.1722308	−1.0	1.312	10.0
6	26.62	1.5	0.4375	3.75	0.00784	1.69504293	3.0	2.663	0.51
7	17.28	1.75	0.6062	4.375	0.008979	1.82395972	4.5	1.819	0.4808
8	25.53	4.75	0.8313	3.875	0.006004	1.42794407	2.5	1.144	17.4
9	17.16	2.25	0.2688	4.125	0.008979	1.01795159	3.5	0.8063	2.869
10	10.16	3.25	0.4938	3.625	0.02088	1.13577944	5.5	2.831	1.35
11	30.35	2.625	0.3531	3.938	0.01428	1.46377777	−0.25	0.3844	6.132
12	11.85	4.625	0.8031	4.938	0.01596	1.08695585	3.75	1.734	0.825
13	11.48	1.625	0.5781	3.438	0.01624	1.23413824	5.75	1.059	0.2112
14	15.04	3.125	0.4656	4.688	0.008102	1.64002595	4.75	0.7219	3.852
15	19.79	2.125	0.2406	3.188	0.008939	1.33435987	2.75	2.747	2.585
16	45.66	4.375	0.4094	3.562	0.01817	1.05772043	1.25	0.5531	41.14
17	46.16	2.375	0.8594	4.562	0.004406	1.39461942	5.25	1.903	13.07
18	50.41	3.375	0.6344	3.062	0.008497	1.75599626	3.25	1.228	21.78
19	31.73	3.875	0.9719	3.812	0.007288	1.54445643	−1.75	2.241	10.76
20	10.6	2.875	0.7469	4.312	0.01417	1.28134533	0.25	1.566	0.6607
21	20.65	2.812	0.4797	3.594	0.007336	1.34869001	−0.625	1.692	5.986

TABLE 2. For caption see on next page.

Index	ω_{T_e}	η_e	τ	q	β_e	$\lambda_D = \lambda_{De}/\rho_e$	$\hat{s}$	Z_{eff}	Q_e/Q_{GB}
22	10.84	3.812	0.2547	4.094	0.01115	1.66685386	1.375	1.017	4.193
23	21.47	1.812	0.7047	3.094	0.01422	1.16278488	5.375	2.367	0.9622
24	12.59	1.312	0.8172	3.844	0.01807	1.24544945	0.375	0.6797	0.1328
25	32.59	2.312	0.1422	3.344	0.01008	1.48275739	2.375	1.355	6.441
26	21.97	4.312	0.5922	4.344	0.008684	1.09465979	−1.625	2.705	9.052
27	12.83	2.062	0.9859	4.969	0.01587	1.29399103	5.875	0.5109	0.1904
28	61.83	4.062	0.5359	3.969	0.01447	1.01174754	1.875	1.861	78.36
29	32.83	3.062	0.7609	3.469	0.003641	1.56675126	3.875	2.536	21.71
30	46.2	3.562	0.8734	3.719	0.007383	1.78897427	−1.125	0.8484	17.02
31	6.201	1.562	0.4234	4.719	0.00822	1.20199048	2.875	2.198	0.2105
32	38.64	4.562	0.1984	3.219	0.02602	1.06480002	4.875	1.523	32.98
33	44.76	2.562	0.6484	4.219	0.008917	1.4109689	0.875	2.873	11.04
34	23.31	3.438	0.6766	4.406	0.01075	1.14457745	2.625	0.4266	6.115
35	5.56	2.438	0.9016	3.906	0.01902	1.02427119	4.625	2.452	0.2475
36	52.97	2.938	0.5641	3.156	0.009562	1.07942823	1.625	0.7641	34.79
37	8.779	1.938	0.3391	3.656	0.004913	1.8610814	3.625	1.439	0.8462
38	24.73	1.688	0.7328	4.031	0.005436	1.52304234	4.125	0.5953	1.088
39	27.29	2.688	0.5078	4.531	0.0185	0.99968494	−1.875	1.27	2.196
40	14.26	4.188	0.6203	3.281	0.008179	1.37879224	−0.875	0.9328	8.305

TABLE 2. Summary of the main parameters (columns two to nine, in the same units as in the main text) and of the corresponding GB-normalized electron heat flux (last column) for the 40 out-of-bounds testing parameters used to validate our proposed surrogate model.

Appendix D. Unit conversion

For simplicity, in the following we use superscript Fa22 to refer to quantities used in our original UQ study in Farcaş *et al.* (2022) and Fa24 to refer to quantities used in the present paper. Analogously, superscripts Ha22 and Ha24 are employed to refer to quantities specific to Hatch *et al.* (2022) and Hatch *et al.* (2024), respectively.

In Farcaş *et al.* (2022), our simulations adopted the generalized Miller geometry using $R = 1.68$ m, with the temperature gradient defined as

$$\omega_{T_e}^{\text{Fa22}} = \frac{R}{T_e} \frac{\mathrm{d}T_e}{\mathrm{d}r}. \tag{D1}$$

The conversion factor between $\omega_{T_e}^{\text{Fa22}}$ in (D1) and $\omega_{T_e}^{\text{Fa24}}$ used in this paper is 4.6576, therefore $\omega_{T_e}^{\text{Fa24}} = \omega_{T_e}^{\text{Fa22}}/4.5676$.

To use the model in (4.2) in the present work, we must (i) convert the temperature gradient and Debye length in our units to be compatible with the units used in (4.2) and (ii) convert the ensuing flux predictions to account for the area definition. The Debye length conversion is straightforward and was explained in the main text. To convert $\omega_{T_e}^{\text{Fa24}}$ to $\omega_{T_e}^{\text{Ha24}}$, we can use the chain rule to obtain

$$\omega_{T_e}^{\text{Ha24}} = \frac{1}{T_e} \frac{\mathrm{d}T_e}{\mathrm{d}\psi_N} = \frac{R}{T_e} \frac{\mathrm{d}T_e}{\mathrm{d}r} \frac{1}{R} \frac{\mathrm{d}r}{\mathrm{d}\psi_N} = \omega_{T_e}^{\text{Fa22}} \frac{\psi_N}{R} \frac{\mathrm{d}r}{\mathrm{d}\psi}$$

$$= 4.5676 \times \omega_{T_e}^{\text{Fa24}} \frac{\psi_N}{R} \frac{\mathrm{d}r}{\mathrm{d}\psi}. \tag{D2}$$

Using the definition of the Miller geometry, we can evaluate $dr/d\psi$ in (D2) for any set of input parameters. The same is not true for ψ_N/R, which requires the knowledge of a global equilibrium. The value of this factor for our nominal parameters is 1/0.31, however, since the results of (4.2) linearly scale with it, it is clear that a model comparison is inevitably skewed by the specific choice of this factor. Since the model in (4.2) is richer than (4.1), we nevertheless employ it in § 4 for a more comprehensive overview using a constant value 1/0.20, which is close to the one for our nominal inputs.

The predictions $Q_e^{\text{Ha24}}/Q_{\text{GB}}^{\text{Ha24}}$ provided by (4.2) can be converted to $Q_e^{\text{Fa24}}/Q_{\text{GB}}^{\text{Fa24}}$ as

$$Q_e^{\text{Fa24}}/Q_{\text{GB}}^{\text{Fa24}} = Q_e^{\text{Ha24}}/Q_{\text{GB}}^{\text{Ha24}} A^{\text{Ha24}}/V'^{\text{Ha22}}, \tag{D3}$$

where $A^{\text{Ha24}} = 51.07$ m^2 and $V'^{\text{Ha22}} = 37.28$ m^2. Moreover, $Q_{\text{GB}}^{\text{Ha24}} = Q_{\text{GB}}^{\text{Fa24}} = Q_{\text{GB}}^{\text{Ha22}}$.

REFERENCES

BELLI, E.A., CANDY, J. & SFILIGOI, I. 2022 Spectral transition of multiscale turbulence in the tokamak pedestal. *Plasma Phys. Control. Fusion* **65** (2), 024001.

BELLI, E.A., CANDY, J. & SFILIGOI, I. 2024 Flow-shear destabilization of multiscale electron turbulence. *Plasma Phys. Control. Fusion* **66** (4), 045019.

CANDY, J. 2009 A unified method for operator evaluation in local Grad–Shafranov plasma equilibria. *Plasma Phys. Control. Fusion* **51** (10), 105009.

CHAPMAN-OPLOPOIOU, B., HATCH, D.R., FIELD, A.R., FRASSINETTI, L., HILLESHEIM, J.C., HORVATH, L., MAGGI, C.F., PARISI, J.F., ROACH, C.M., SAARELMA, S., WALKER, J. & JET CONTRIBUTORS 2022 The role of ETG modes in JET–ILW pedestals with varying levels of power and fuelling. *Nucl. Fusion* **62** (8), 086028.

DORLAND, W., JENKO, F., KOTSCHENREUTHER, M. & ROGERS, B.N. 2000 Electron temperature gradient turbulence. *Phys. Rev. Lett.* **85**, 5579–5582.

EFRON, B. & TIBSHIRANI, R. 1986 Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* **1** (1), 54–75.

FARCAŞ, I.-G., MERLO, G. & JENKO, F. 2022 A general framework for quantifying uncertainty at scale. *Commun. Engng* **1** (1), 43.

GUTTENFELDER, W., GROEBNER, R.J., CANIK, J.M., GRIERSON, B.A., BELLI, E.A. & CANDY, J. 2021 Testing predictions of electron scale turbulent pedestal transport in two DIII-D ELMy H-modes. *Nucl. Fusion* **61** (5), 056005.

HASSAN, E., HATCH, D.R., HALFMOON, M.R., CURIE, M., KOTCHENREUTHER, M.T., MAHAJAN, S.M., MERLO, G., GROEBNER, R.J., NELSON, A.O. & DIALLO, A. 2021 Identifying the microtearing modes in the pedestal of DIII-D H-modes using gyrokinetic simulations. *Nucl. Fusion* **62** (2), 026008.

HATCH, D.R., KOTSCHENREUTHER, M.T., LI, P.-Y., CHAPMAN-OPLOPOIOU, B., PARISI, J., MAHAJAN, S.M. & GROEBNER, R. 2024 Modeling electron temperature profiles in the pedestal with simple formulas for ETG transport. *Nucl. Fusion* **64** (6), 066007.

HATCH, D.R., KOTSCHENREUTHER, M.T., MAHAJAN, S., VALANJU, P. & LIU, X. 2017 A gyrokinetic perspective on the JET-ILW pedestal. *Nucl. Fusion* **57** (3), 036020.

HATCH, D.R., MICHOSKI, C., KUANG, D., CHAPMAN-OPLOPOIOU, B., CURIE, M., HALFMOON, M., HASSAN, E., KOTSCHENREUTHER, M., MAHAJAN, S.M., MERLO, G., PUESCHEL, M.J., WALKER, J. & STEPHENS, C.D. 2022 Reduced models for ETG transport in the tokamak pedestal. *Phys. Plasmas* **29** (6), 062501.

HATCH, D.R., TOLD, D., JENKO, F., DOERK, H., DUNNE, M.G., WOLFRUM, E., VIEZZER, E., THE ASDEX UPGRADE TEAM & PUESCHEL, M.J. 2015 Gyrokinetic study of ASDEX Upgrade inter-ELM pedestal profile evolution. *Nucl. Fusion* **55** (6), 063028.

IDOMURA, Y. 2006 Self-organization in electron temperature gradient driven turbulence. *Phys. Plasmas* **13** (8), 080701.

JENKO, F. & DORLAND, W. 2002 Prediction of significant tokamak turbulence at electron gyroradius scales. *Phys. Rev. Lett.* **89**, 225001.

JENKO, F., DORLAND, W. & HAMMETT, G.W. 2001 Critical gradient formula for toroidal electron temperature gradient modes. *Phys. Plasmas* **8** (9), 4096–4104.

JENKO, F., DORLAND, W., KOTSCHENREUTHER, M. & ROGERS, B.N. 2000 Electron temperature gradient driven turbulence. *Phys. Plasmas* **7** (5), 1904–1910.

JENKO, F., TOLD, D., XANTHOPOULOS, P., MERZ, F. & HORTON, L.D. 2009 Gyrokinetic turbulence under near-separatrix or nonaxisymmetric conditions. *Phys. Plasmas* **16** (5), 055901.

LEPPIN, L.A., GÖRLER, T., CAVEDON, M., DUNNE, M.G., WOLFRUM, E. & JENKO, F. 2023 Complex structure of turbulence across the ASDEX Upgrade pedestal. *J. Plasma Phys.* **89** (6), 905890605.

LI, P.-Y., HATCH, D.R., CHAPMAN-OPLOPOIOU, B., SAARELMA, S., ROACH, C.M., Kotschenreuther, M., Mahajan, S.M., Merlo, G. & the MAST Team 2023 ETG turbulent transport in the Mega Ampere Spherical Tokamak (MAST) pedestal. *Nucl. Fusion* **64** (1), 016040.

PARISI, J.F., *et al.* 2020 Toroidal and slab ETG instability dominance in the linear spectrum of JET-ILW pedestals. *Nucl. Fusion* **60** (12), 126045.

PARISI, J.F., *et al.* 2022 Three-dimensional inhomogeneity of electron-temperature-gradient turbulence in the edge of tokamak plasmas. *Nucl. Fusion* **62** (8), 086045.

STIMMEL, K., GIL, L., GÖRLER, T., CAVEDON, M., DAVID, P., DUNNE, M., DUX, R., FISCHER, R., JENKO, F., KALLENBACH, A., *et al.* 2022 Gyrokinetic analysis of an argon-seeded EDA H-mode in ASDEX Upgrade. *J. Plasma Phys.* **88** (3), 905880315.

TOLD, D., JENKO, F., XANTHOPOULOS, P., HORTON, L.D., WOLFRUM, E. & ASDEX UPGRADE TEAM 2008 Gyrokinetic microinstabilities in ASDEX Upgrade edge plasmas. *Phys. Plasmas* **15** (10), 102306.

TREFETHEN, L.N. 2012 *Approximation Theory and Approximation Practice (Applied Mathematics)*. Society for Industrial and Applied Mathematics.

WALKER, J. & HATCH, D.R. 2023 ETG turbulence in a tokamak pedestal. *Phys. Plasmas* **30** (8), 082307.