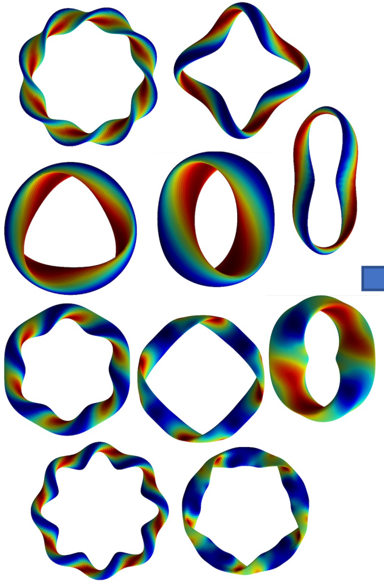
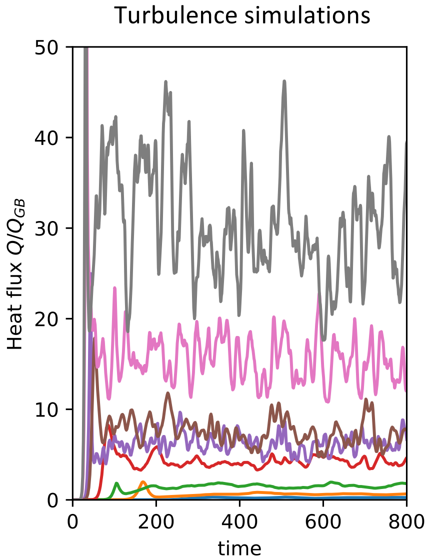
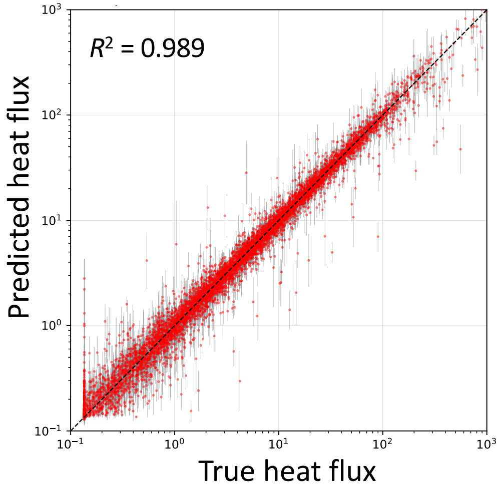
Efficient interpretable regression and classification with translation-invariance





Regression, classification



Feature importance

$$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$$

M Landreman, J Y Choi, C Alves, P Balaprakash, R M Churchill , R Conlin, G Roberg-Clark

Journal of Plasma Physics **91**, E120 (2025), arXiv:2502.11657

Efficient interpretable regression and classification with translation-invariance

- Motivating application: turbulence in fusion plasmas
- Regression/classification method
- Results

Efficient interpretable regression and classification with translation-invariance

- Motivating application: turbulence in fusion plasmas
- Regression/classification method
- Results

THE WALL STREET JOURNAL.

Nuclear-Fusion Startup Lands $1.8 Billion as Investors Chase Star Power

Proxima Fusion

Home Technology About Careers Press & News

Proxima Fusion Raises €130M Series A to Build World's First Stellarator-Based Fusion Power Plant in 2030s

CLIMATE

Google inks its first fusion power deal with Commonwealth Fusion Systems

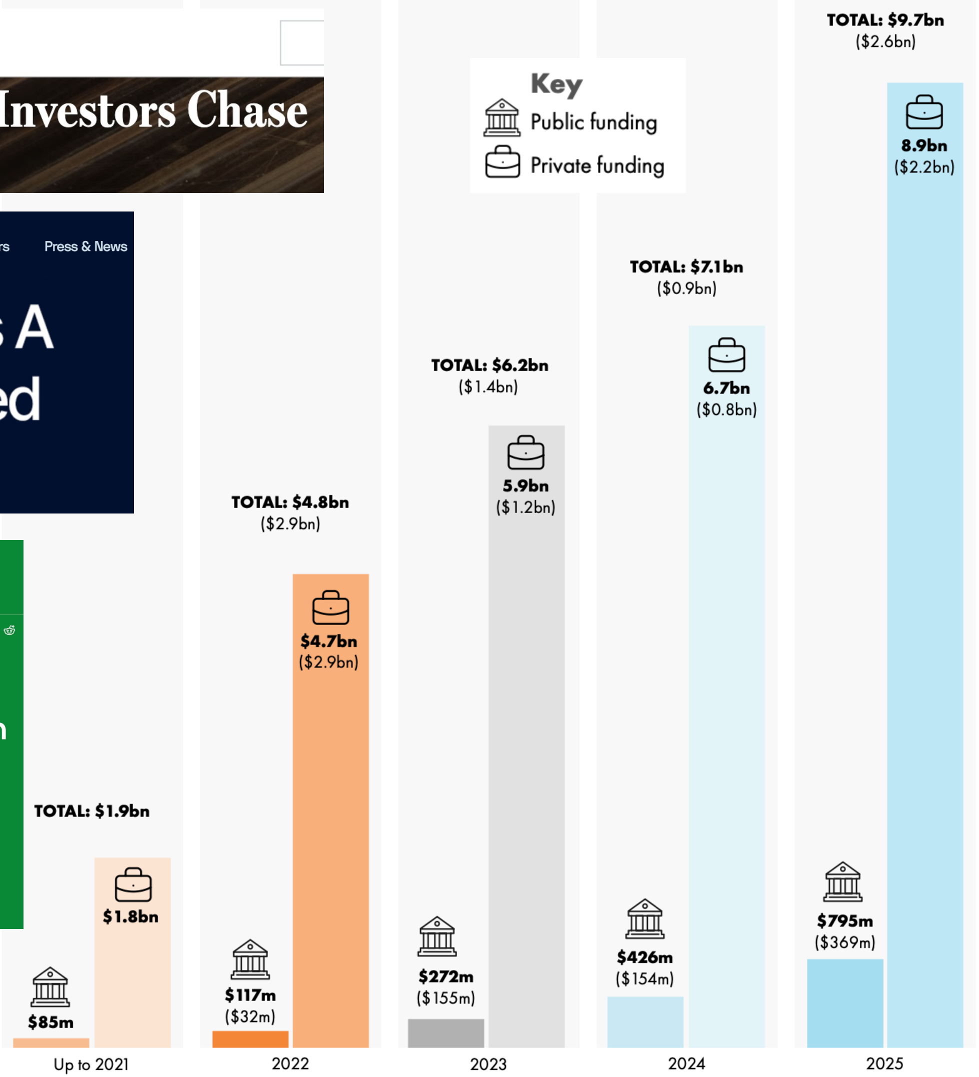
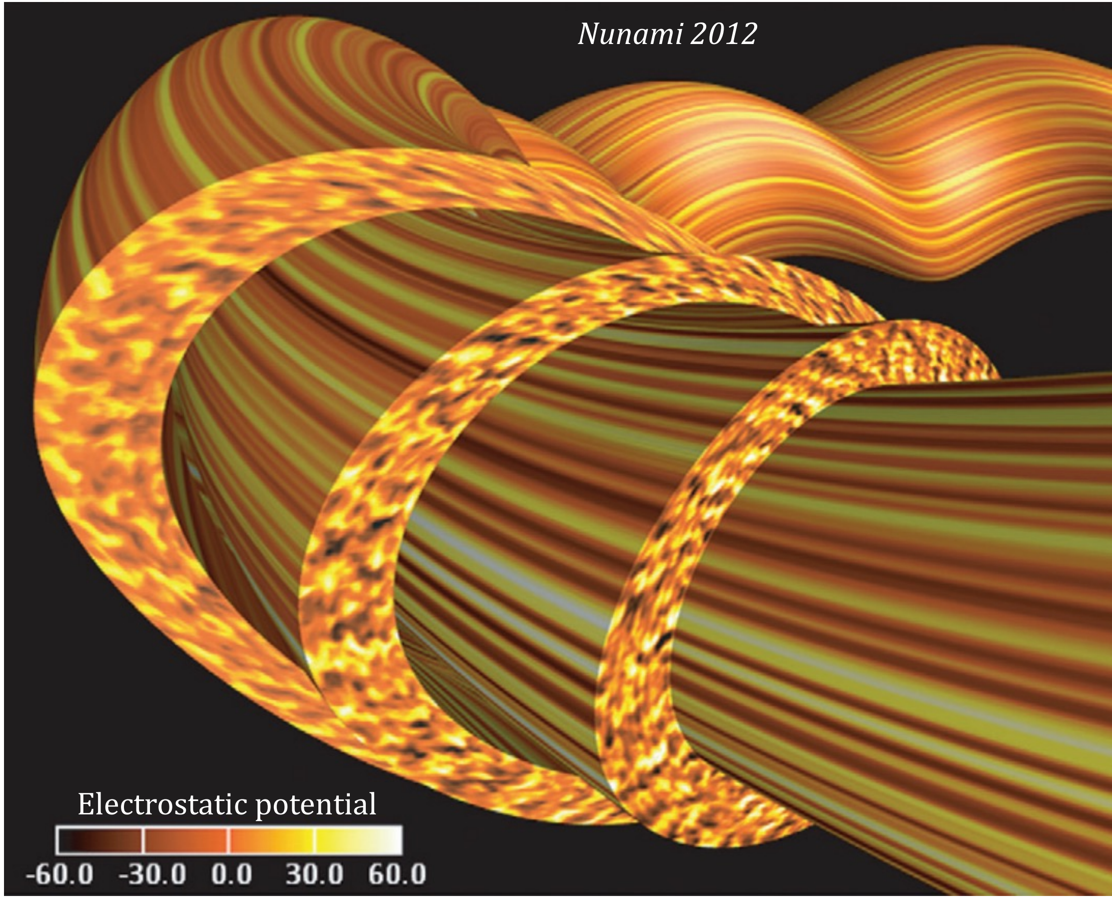


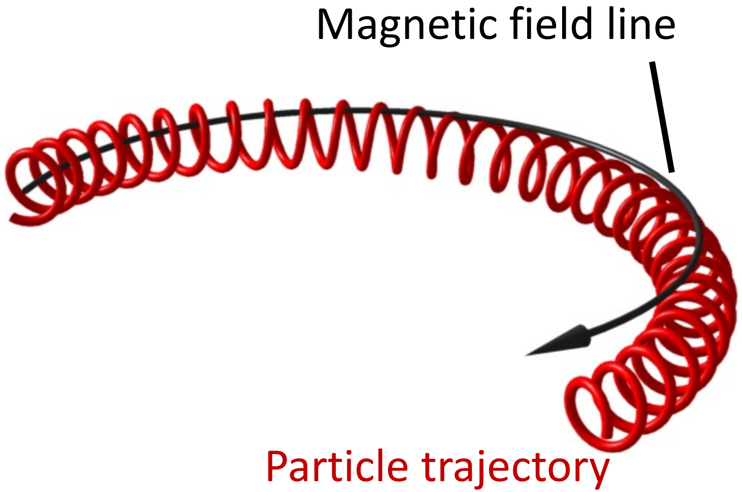
Chart from Fusion Industry Association

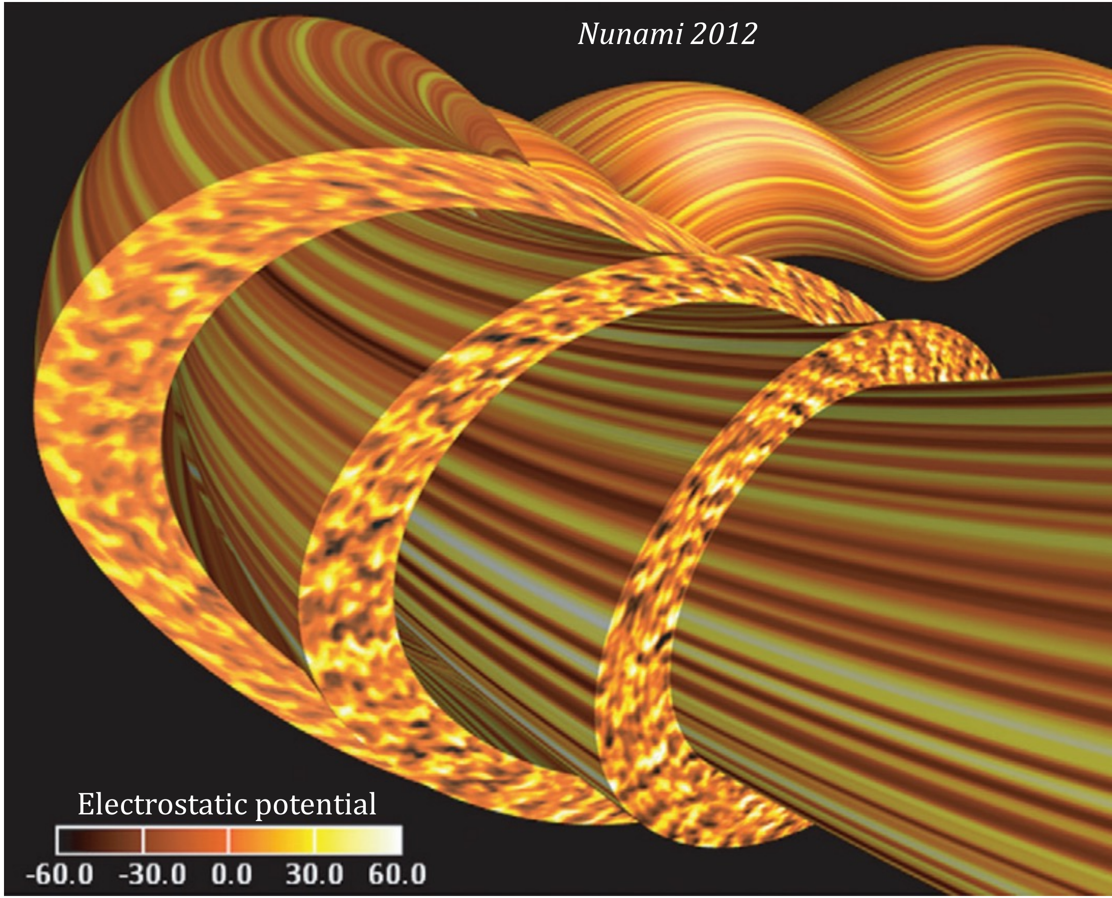
Turbulence in fusion plasmas: we want to understand & optimize

- Gradients of plasma density & temperature cause instabilities & turbulence.
- Causes heat to leak out.
- Known to depend on plasma shape, but how exactly?
- Geometry could be optimized to reduce turbulence
- Turbulence simulations: ~10 GPU-min to 10^6s of CPU-hours.
- 5D + time



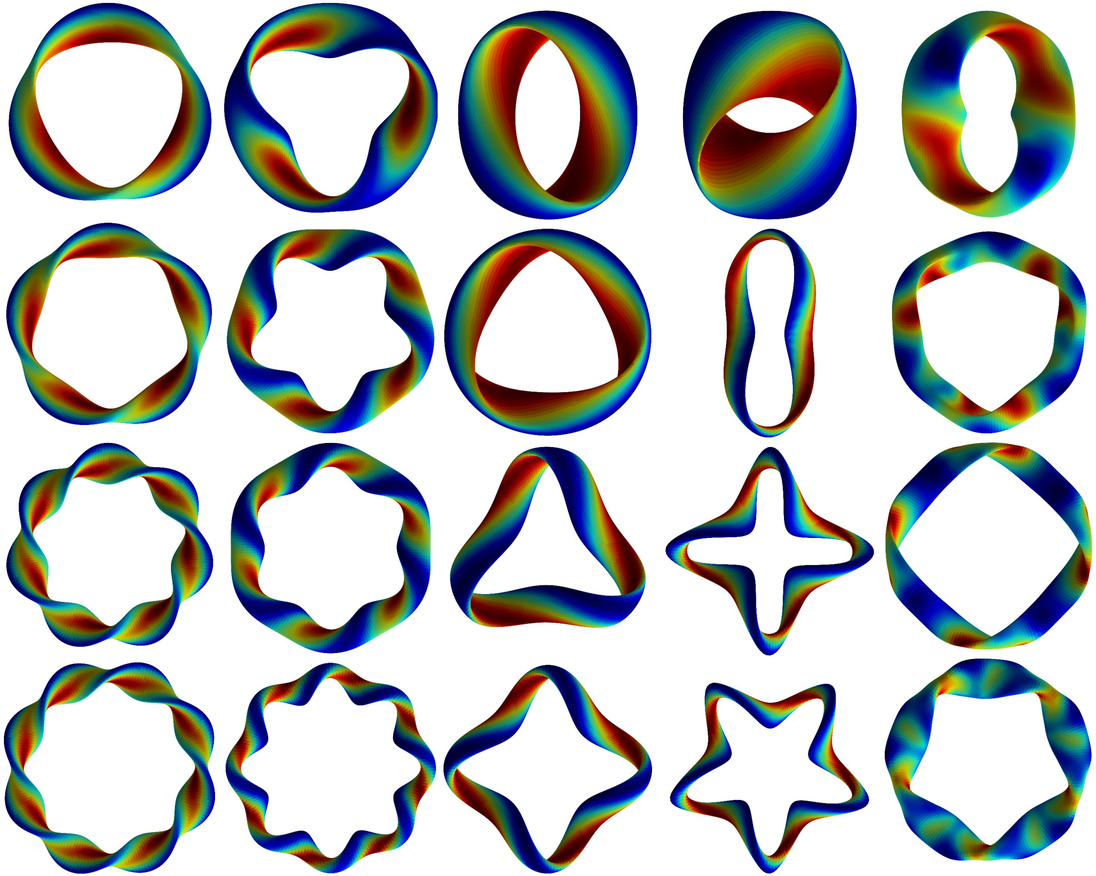
Turbulence in fusion plasmas is highly anisotropic due to magnetic field

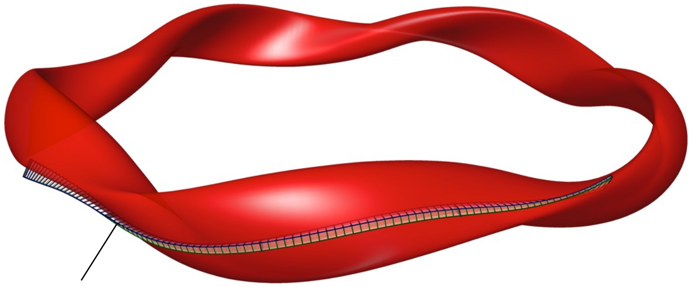




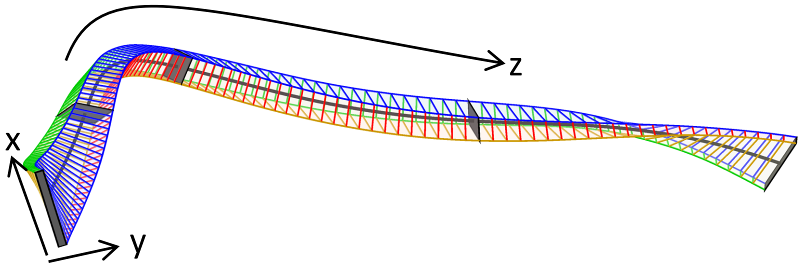
Approach here: learn how turbulence depends on geometry using data

Start with direct numerical simulations of turbulence in 10^5 geometries





“Flux tube” simulation domain



Equations describing the turbulence

Electrostatic potential: $$\Phi(x,y,z,t) = \sum_{\mathbf{k}} \hat{\Phi}_{\mathbf{k}}(z,t)\exp(ik_x x + ik_y y) \qquad \mathbf{k} = (k_x, k_y)$$

Ion distribution function: $$h(x,y,z,v_{||},\mu,t) = \sum_{\mathbf{k}} \hat{h}_{\mathbf{k}}(z,v_{||},\mu,t)\exp(ik_x x + ik_y y) \qquad \mu = \frac{v_\perp^2}{2B}$$

Gyrokinetic equation: $$\frac{\partial \hat{h}_{\mathbf{k}}}{\partial t} + v_{||}\frac{\partial \hat{h}_{\mathbf{k}}}{\partial z} - \mu \frac{\partial B}{\partial z}\frac{\partial \hat{h}_{\mathbf{k}}}{\partial v_{||}} + \mathbf{v}_d \cdot \nabla \hat{h}_{\mathbf{k}} + \mathcal{N}_{\mathbf{k}} = \frac{eJ_{0,\mathbf{k}}}{T}\left(\frac{\partial \hat{\Phi}_{\mathbf{k}}}{\partial t} + i\omega_*^T \hat{\Phi}_{\mathbf{k}}\right)$$

$$\mathbf{v}_d = \frac{mv_\perp^2}{2eB^3}\mathbf{B}\times\nabla B + \frac{mv_{||}^2}{eB^2}\mathbf{B}\times(\mathbf{b}\cdot\nabla\mathbf{b}) \qquad \mathbf{b} = \frac{1}{B}\mathbf{B} \qquad \mathcal{N}_{\mathbf{k}} = \sum_{\mathbf{k}'} \frac{J_{0,\mathbf{k}'}\hat{\Phi}_{\mathbf{k}'}\hat{h}_{\mathbf{k}''}}{B_{\mathrm{ref}}}\left(k_y' k_x'' - k_x' k_y''\right)$$

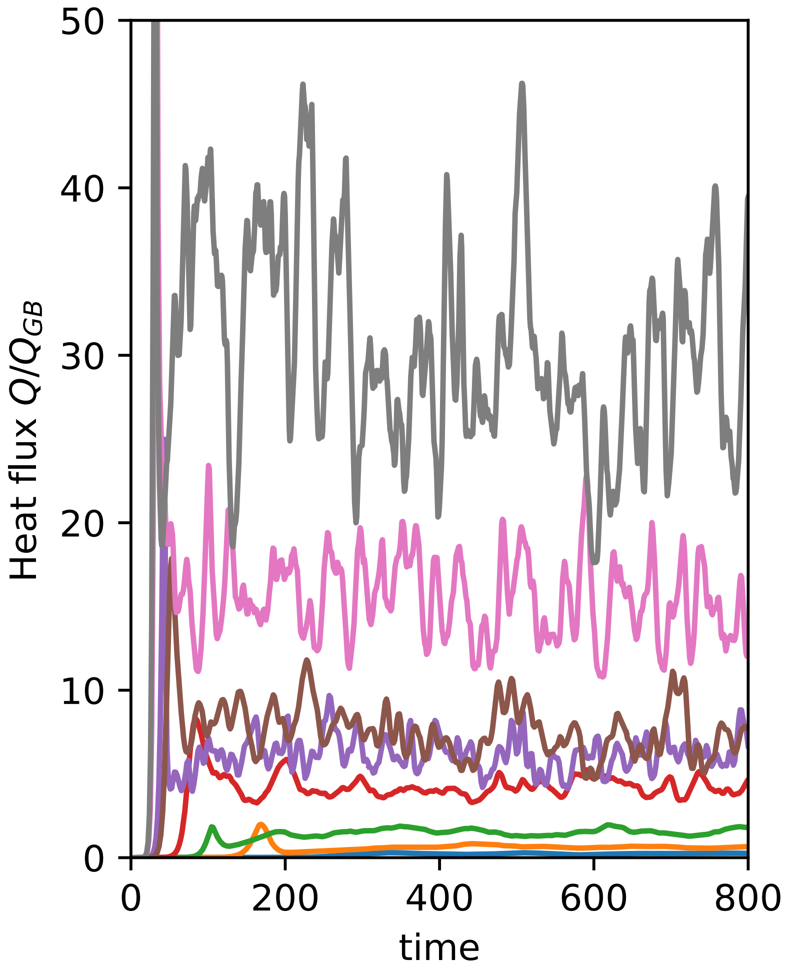
$$\mathbf{k}'' = \mathbf{k} - \mathbf{k}'$$

$$J_{0,\mathbf{k}} = J_0\left(\frac{v_\perp m}{eB}\sqrt{k_x^2\left|\nabla x\right|^2 + 2k_x k_y \nabla x\cdot\nabla y + k_y^2\left|\nabla y\right|^2}\right) \qquad \omega_*^T = \frac{k_y T}{eB_{\mathrm{ref}}}\left[\frac{1}{n}\frac{dn}{dx} + \frac{1}{T}\frac{dT}{dx}\left(\frac{mv^2}{2T} - \frac{3}{2}\right)\right]$$

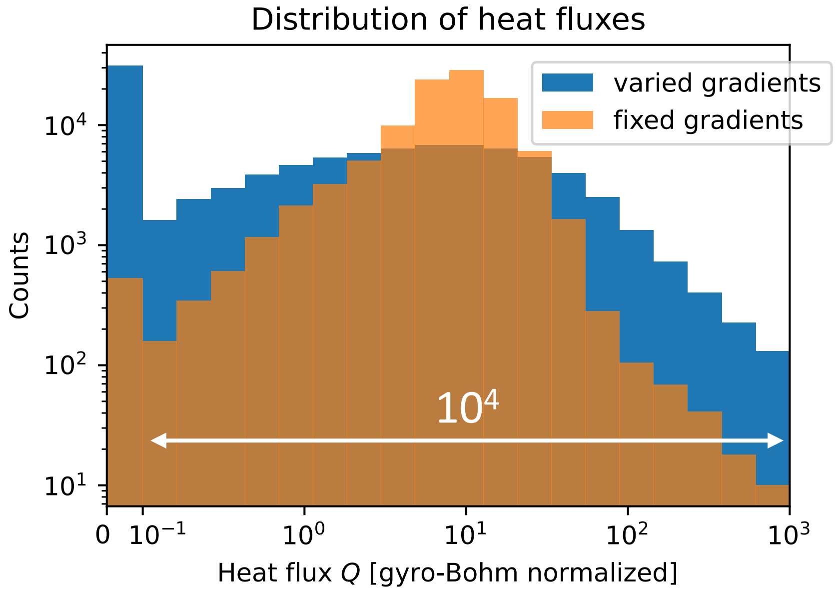
Quasineutrality: $$\hat{\Phi}_{\mathbf{k}}\left(1+\frac{T}{T_e}\right) - \left\langle\hat{\Phi}_{\mathbf{k}}\right\rangle\left(\frac{T}{T_e}\right) = \frac{T}{ne}\int d^3v J_{0,\mathbf{k}}\hat{h}_{\mathbf{k}} \qquad \langle\dots\rangle = \frac{\int \frac{dz}{B}(\dots)}{\int \frac{dz}{B}}$$

Heat flux: $$Q = \sum_{\mathbf{k}}\left\langle \int d^3v \frac{m_i v^2}{2} ik_y J_{0,\mathbf{k}}\hat{h}_{\mathbf{k}}\hat{\Phi}_{-\mathbf{k}}\right\rangle$$

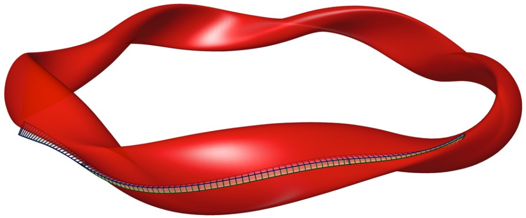
Nonlinear turbulence simulations were run in 10^5 geometries

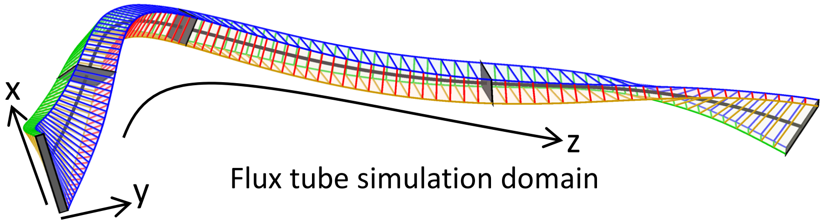


- 1 simulation in each tube with random dT/dx and dn/dx.
- 1 simulation in each tube with (a/T) dT/dx = 3, (a/n) dn/dx=0.9
- 8 minutes to get heat flux on 1 GPU
- 2×10^5 nonlinear simulations took < 7000 node-hours



Raw feature space: 7x 1D functions that enter the turbulence simulations

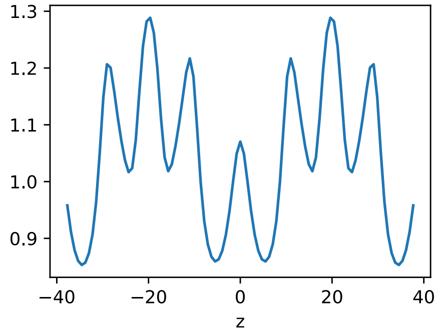




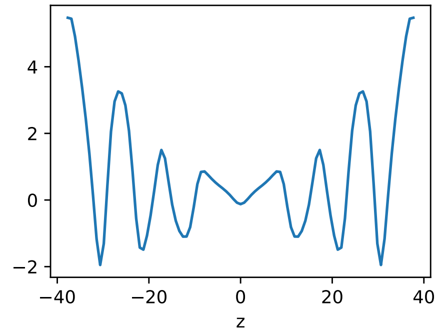
$$\boldsymbol{B} = B_{ref} \nabla x \times \nabla y$$

$$x = a\sqrt{\psi/\psi_{edge}}$$

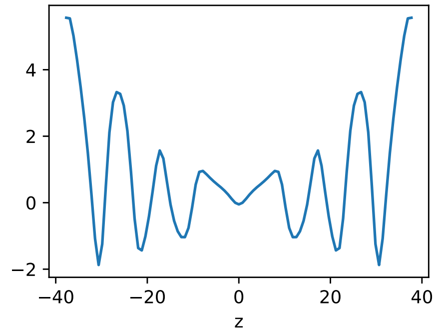
$|B|$



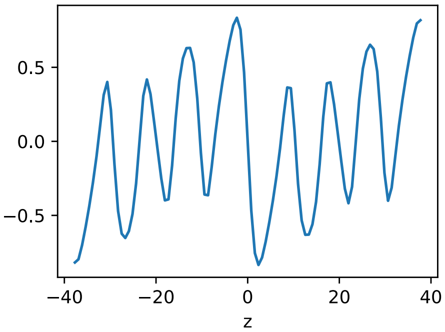
$|B|^{-3}\boldsymbol{B}\times\nabla|B|\cdot\nabla y$



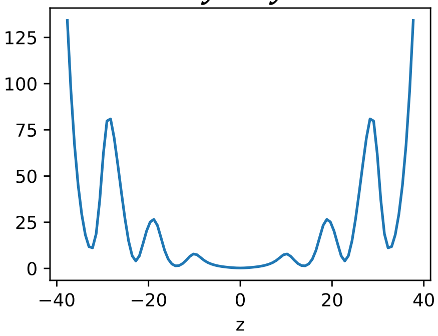
$|B|^{-2}\boldsymbol{B}\times\boldsymbol{\kappa}\cdot\nabla y$



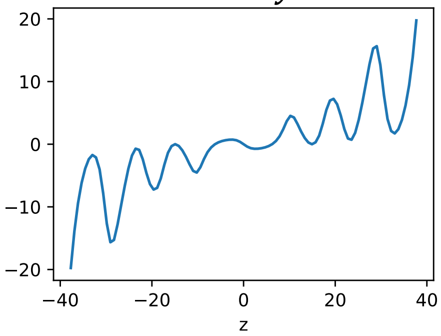
$|B|^{-3}\boldsymbol{B}\times\nabla|B|\cdot\nabla x$



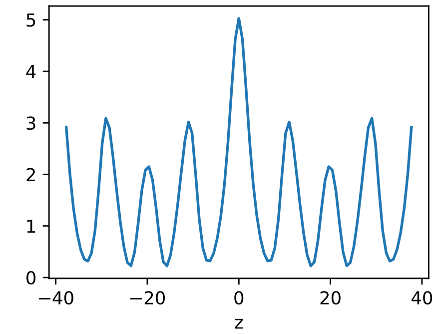
$\nabla y\cdot\nabla y$



$\nabla x\cdot\nabla y$

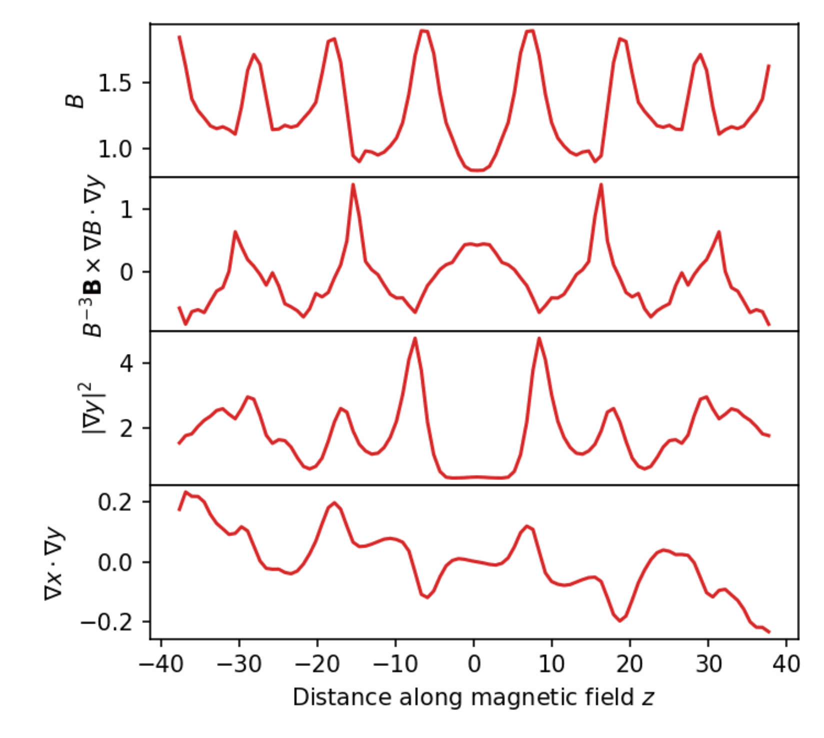


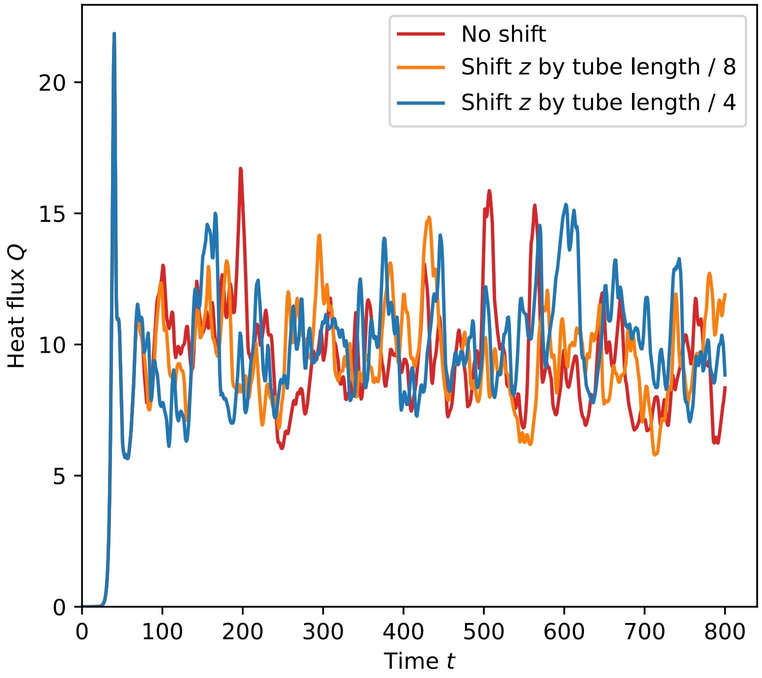
$\nabla x\cdot\nabla x$



Raw features should *not* be directly fed to classical regression or fully-connected neural network, since model should be translation-invariant

- Gyrokinetic equation, hence heat flux, is invariant under periodic translation of the raw features in z.
- Similar to computer vision, where convolutional neural networks give approximate translation-invariance.

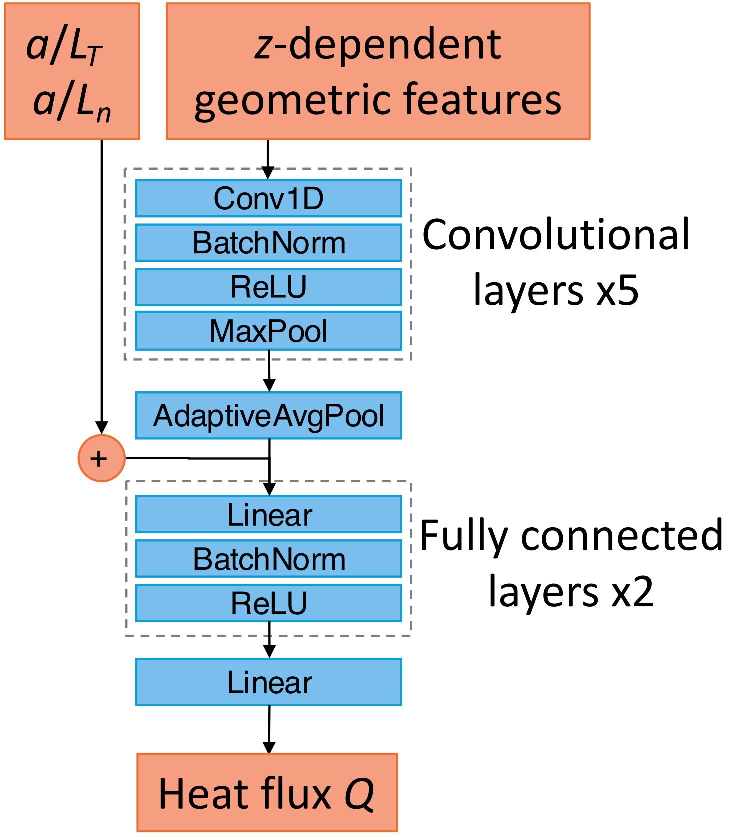


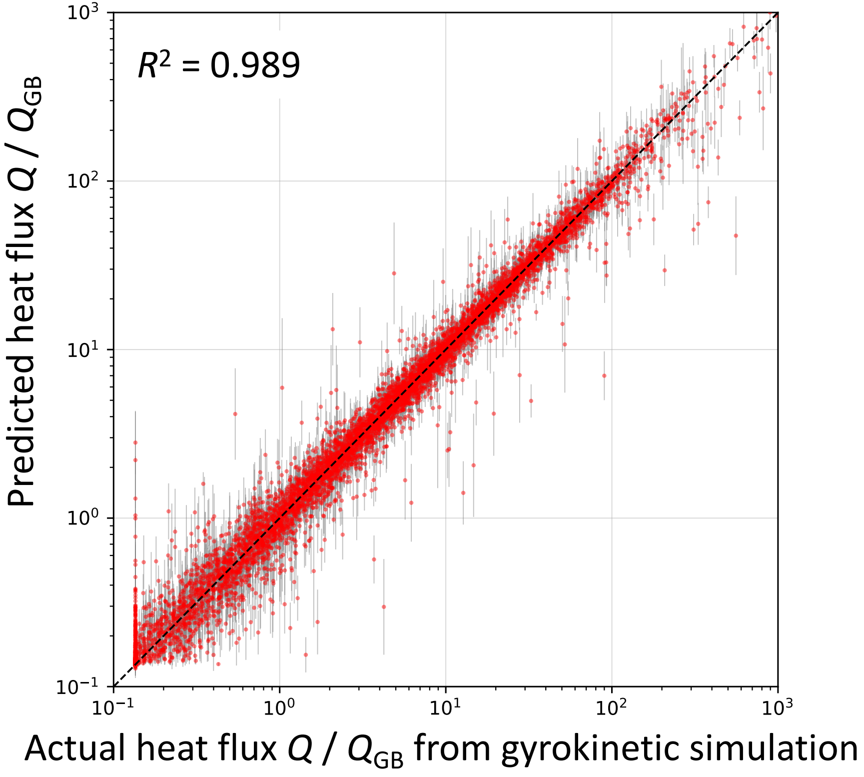


Efficient interpretable regression and classification with translation-invariance

- Motivating application: turbulence in fusion plasmas
- Regression/classification method
- Results

Convolutional neural networks give accurate prediction of the turbulence, but lack interpretability





Regression/classification method to achieve invariance & interpretability

1. Define library of candidate features which are all translation-invariant. (Akin to SINDy).
2. Apply a fast regression/classification method like decision trees to these features.
3. Use forward sequential feature selection to pick out only the most important features.

- Guarantees that model will respect the invariance.
- Each feature is an analytic expression $\Rightarrow$ partially interpretable.
- Step 2 allows extra nonlinearity beyond functions in the library.
- Method in step 2 must be fast to fit because we will fit millions of models.

To create library of candidate features, apply any translation-invariant reduction to translation-equivariant operations

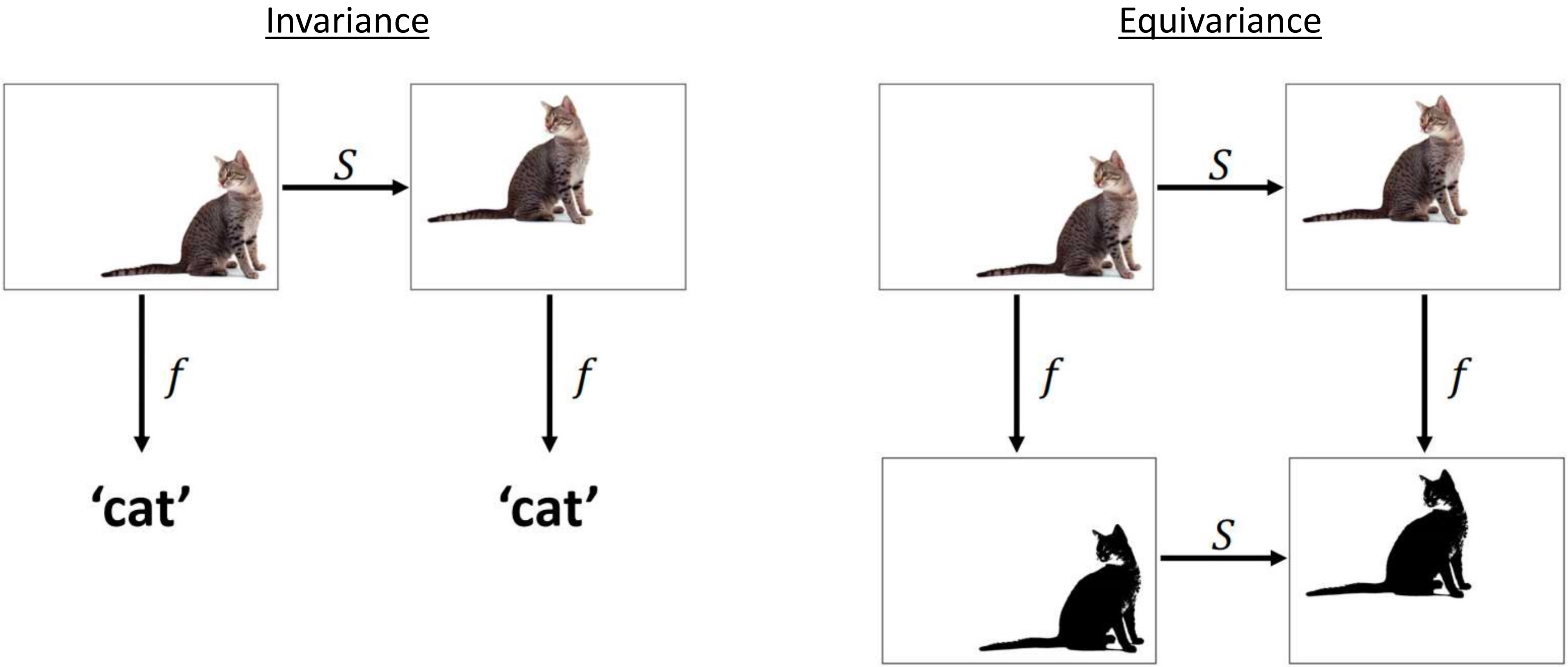


Figure by Bernhard Kainz

To create library of candidate features, apply any translation-invariant reduction to translation-equivariant operations

Start with inputs to the gyrokinetic equation:
$F = \{B,\ B^{-3}\mathbf{B}\times\nabla B\cdot\nabla y,\ B^{-2}\mathbf{B}\times\boldsymbol{\kappa}\cdot\nabla y,\ B^{-3}\mathbf{B}\times\nabla B\cdot\nabla x,\ |\nabla x|^2,\ \nabla x\cdot\nabla y,\ |\nabla y|^2\}$.

U = unary operations on f(z): identity, df/dz, Heaviside(f), Heaviside(-f), ReLU(f), ReLU(-f), 1/f, f^2, f/B (Jacobian), f*B

C(U(F)) = U(F) and all pairwise products of functions in U(F)

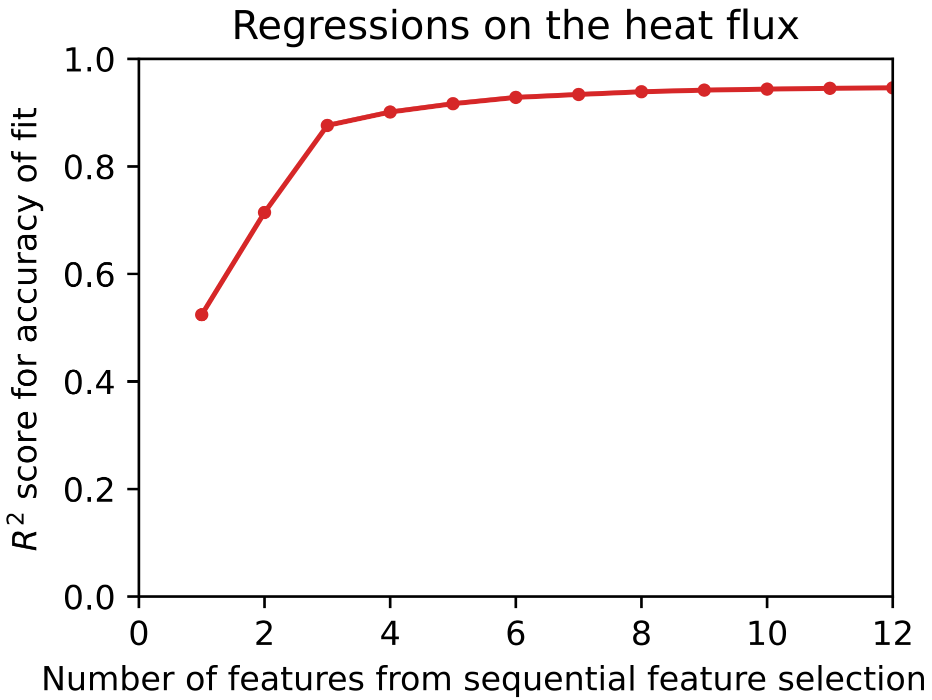
Reductions: R = {min, max, max-min, mean, median, mean square, variance, skewness, L_1 norm, quantiles 0.1, 0.25, 0.75, or 0.9, abs of fft coefficients 1-3, $k_{||}$ with largest amplitude, expected $k_{||}$, count above [-2, -1, 0, 1, 2]}

Features: R(U(C(U(F)))) $\Longrightarrow$ > 1 million combinations

Efficient interpretable regression and classification with translation-invariance

- Motivating application: turbulence in fusion plasmas
- Regression/classification method
- Results

Forward sequential feature selection



Using decision trees (XGBoost library), varied-gradient dataset

Most important features from sequential feature selection

Regression on heat flux

Feature	R^2
a/L_T	0.524
a/L_n	0.714
$\mathrm{mean}(\Theta(\mathbf{B}\times\kappa\cdot\nabla y)\lvert\nabla x\rvert^4/B^2)$	0.876
$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2\lvert\nabla x\rvert^8/B^8)$	0.901
$\mathrm{absFFTCoeff1}(\mathrm{ReLU}(\mathbf{B}\times\nabla B\cdot\nabla x)/B^5)$	0.917

Classification (stability vs instability)

Feature	log-loss
a/L_T	0.361
a/L_n	0.189
$\mathrm{mean}(\Theta(\mathbf{B}\times\nabla B\cdot\nabla y)\lvert\nabla x\rvert^2/B)$	0.122
$\mathrm{mean}(\Theta(-\mathbf{B}\times\nabla B\cdot\nabla x)\lvert\nabla x\rvert^2 B)$	0.105
$\mathrm{mean}((\mathbf{B}\times\kappa\cdot\nabla y)/B)$	0.094

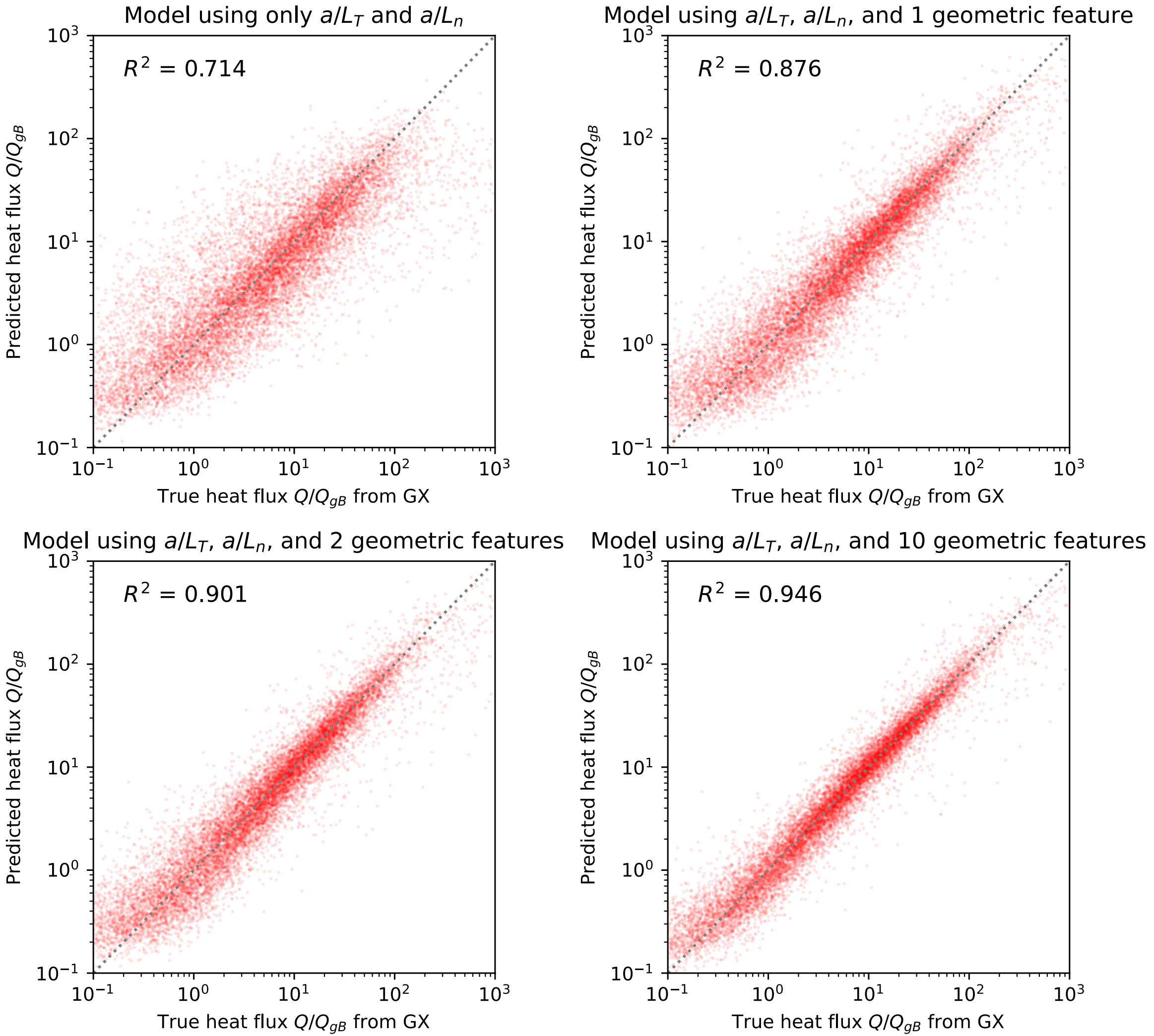
Most important geometric feature:
Local temperature gradient in real space |∇T| = (dT/dx) |∇x|, where there is bad curvature.

The gradients are more important than any geometric feature

Xanthopoulos et al (2011), Nakata & Matsuoka (2022):
Larger geodesic curvature (= radial drift) ⇒ Stronger damping of zonal flows ⇒ higher heat flux

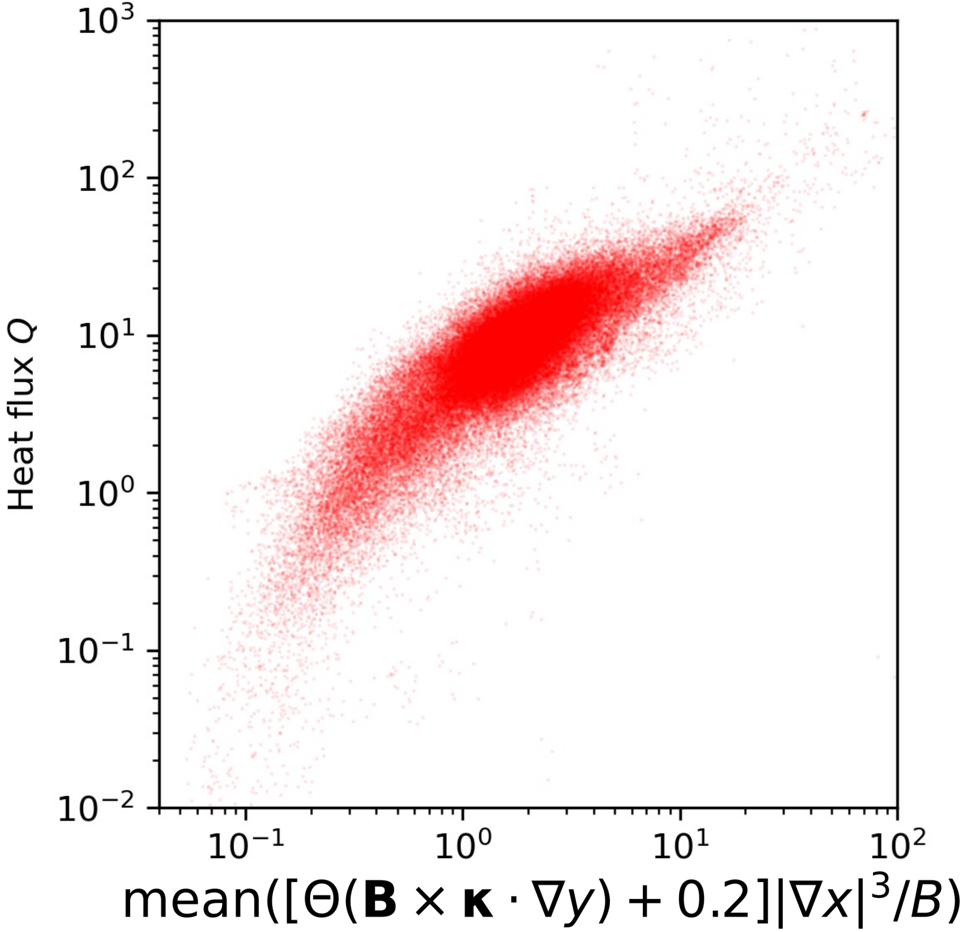
Similar to proposed physics-inspired surrogates
[Mynick (2010), Xanthopoulos (2014), Stroteich (2022), Goodman (2024)].

Sequential feature selection allows closer fit to the data as more geometric features are included



Performance shown on 20% held-out test data

The first geometric feature can be fine-tuned for even better fit

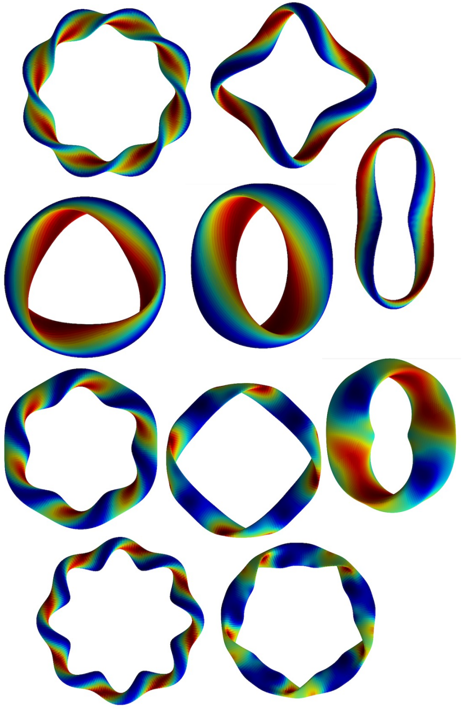


Summary

- Interpretable regression/classification guaranteeing translation-invariance.
- Library of candidate invariant features + decision trees + feature selection.
- Most important features align with recent physics-motivated surrogates.

Future work

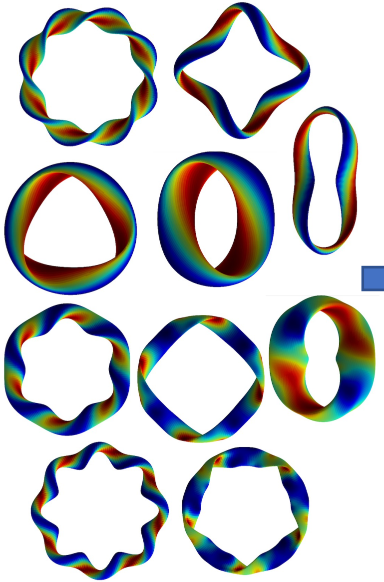
- Use model for optimizing geometry & predicting temperature.
- Compare to saliency maps for the neural networks.
- Include physics-motivated features (e.g. linear growth rates).
- Generate data with higher physics fidelity, repeat analysis.
- Other suggestions for methods to try?

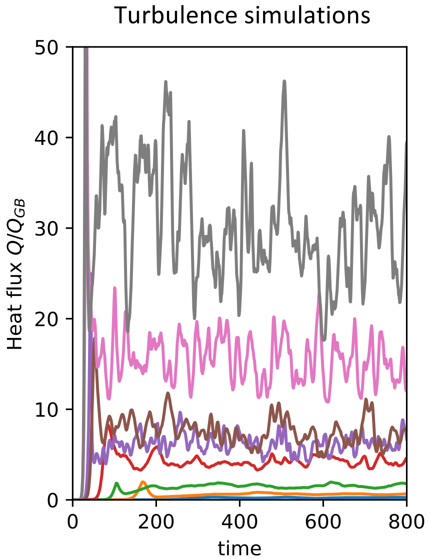


Journal of Plasma Physics 91, E120 (2025)
arXiv:2502.11657

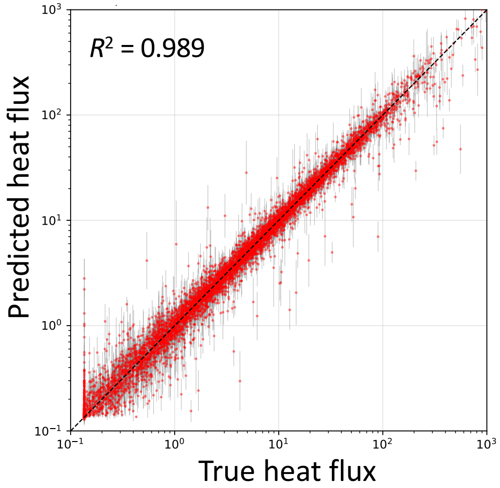
Extra slides

How does magnetic geometry affect ITG turbulence? Insights from data & machine learning





Regression



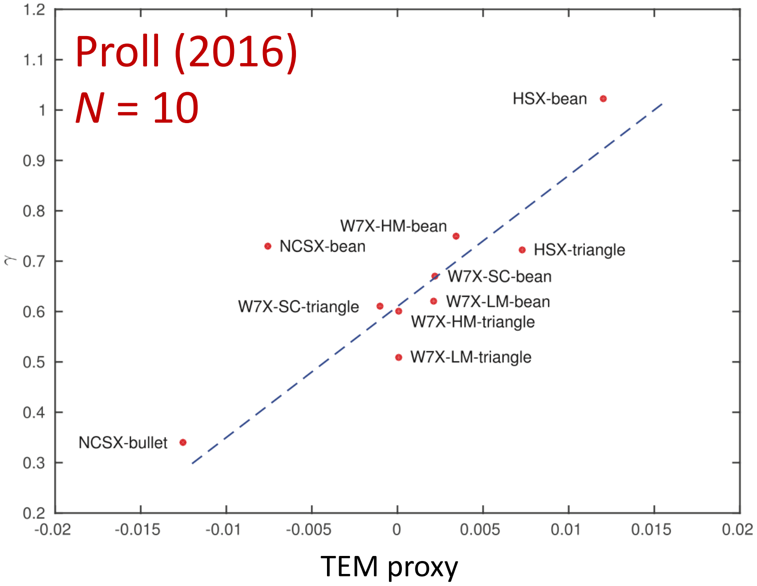
Feature importance

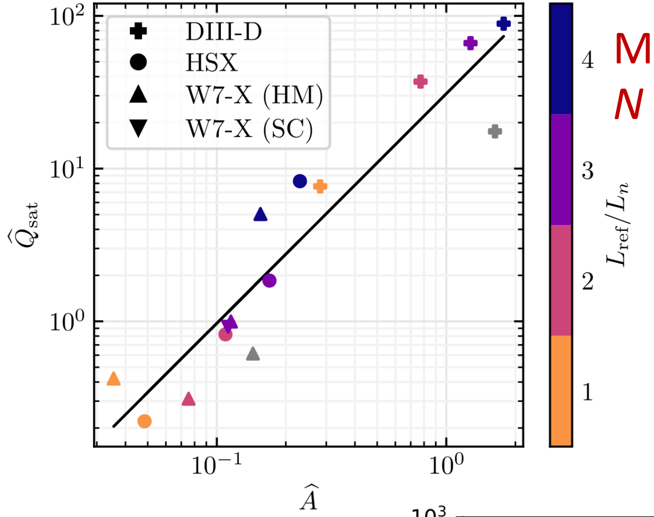
$$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$$
$$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2|\nabla x|^8/B^8)$$

M Landreman, J Y Choi, C Alves, P Balaprakash, R M Churchill , R Conlin, G Roberg-Clark **arXiv:2502.11657**

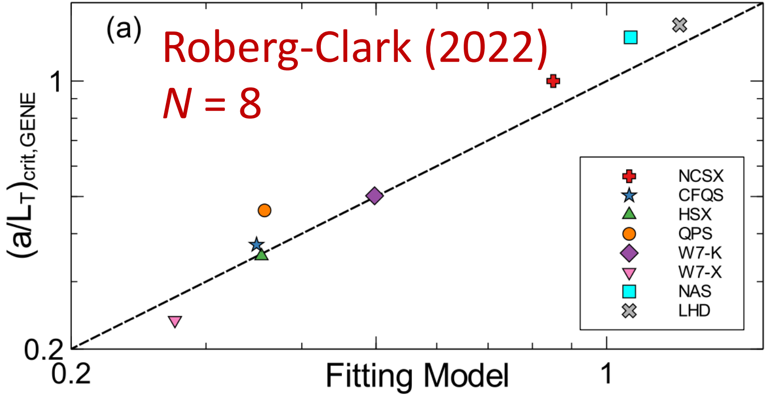
Thanks to many others who gave suggestions

Supported by the US DOE StellFoundry SciDAC

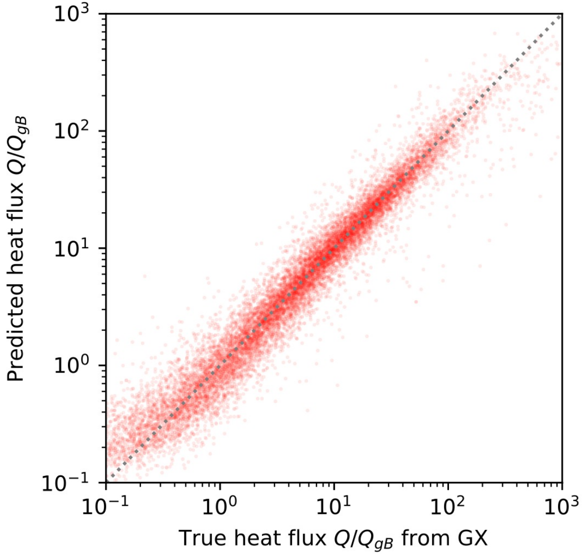




Mackenbach (2022)
N = 15



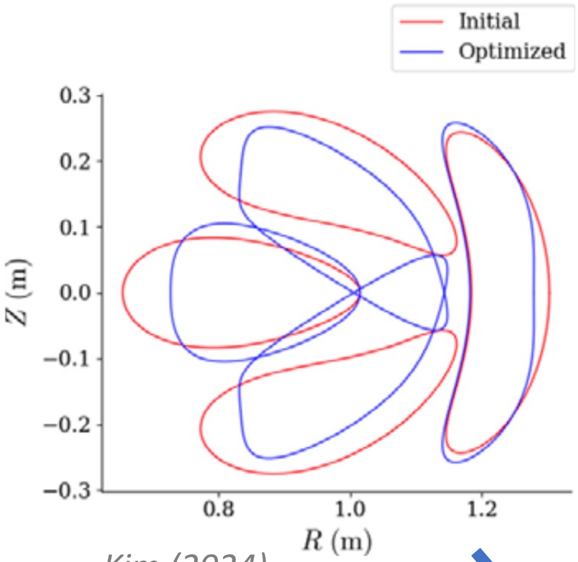
This work:
N = 100,705



Motivations

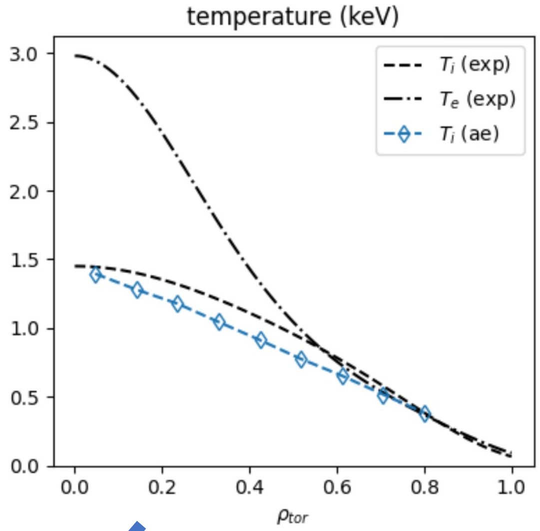
Understanding

Optimization



Kim (2024)

Profile prediction



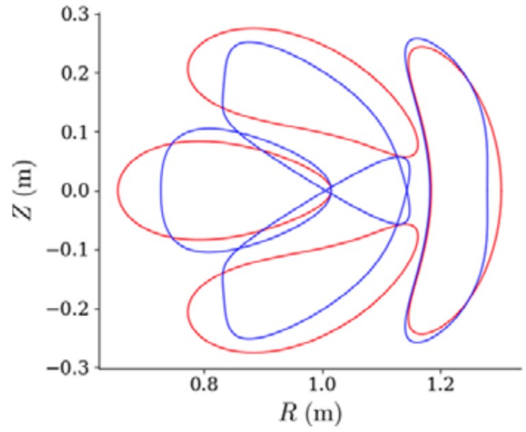
Mandell (2024)

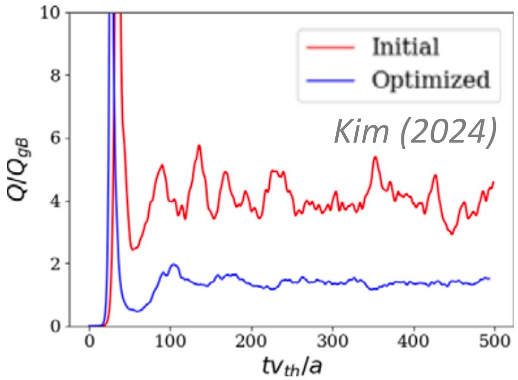
Optimize geometry for maximum fusion power

Motivations

Understanding

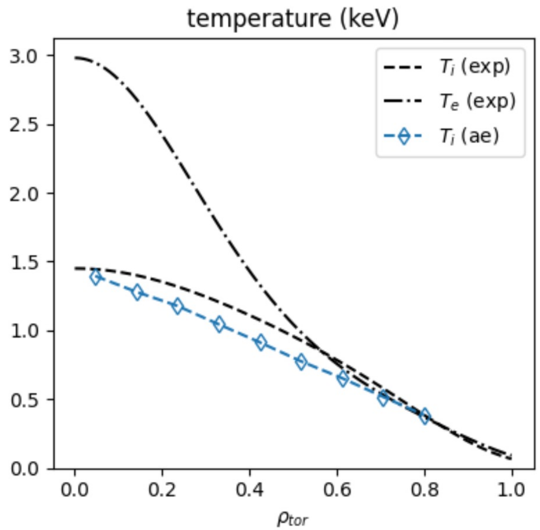
Optimization



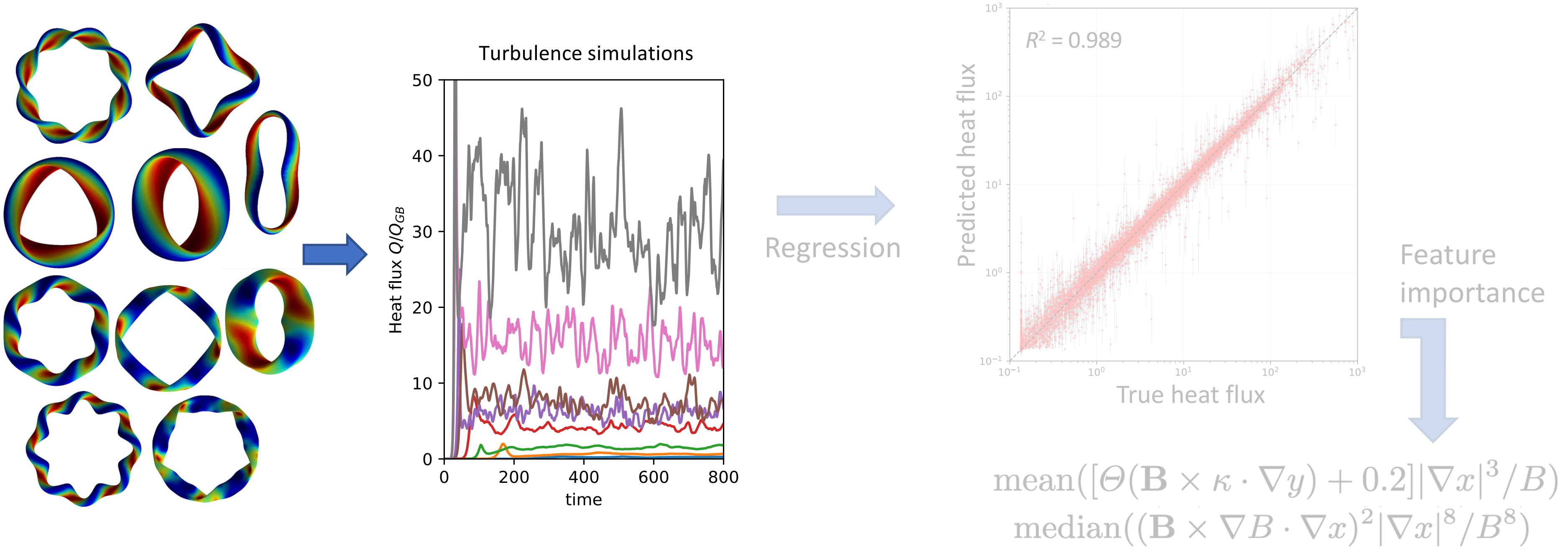


Kim (2024)

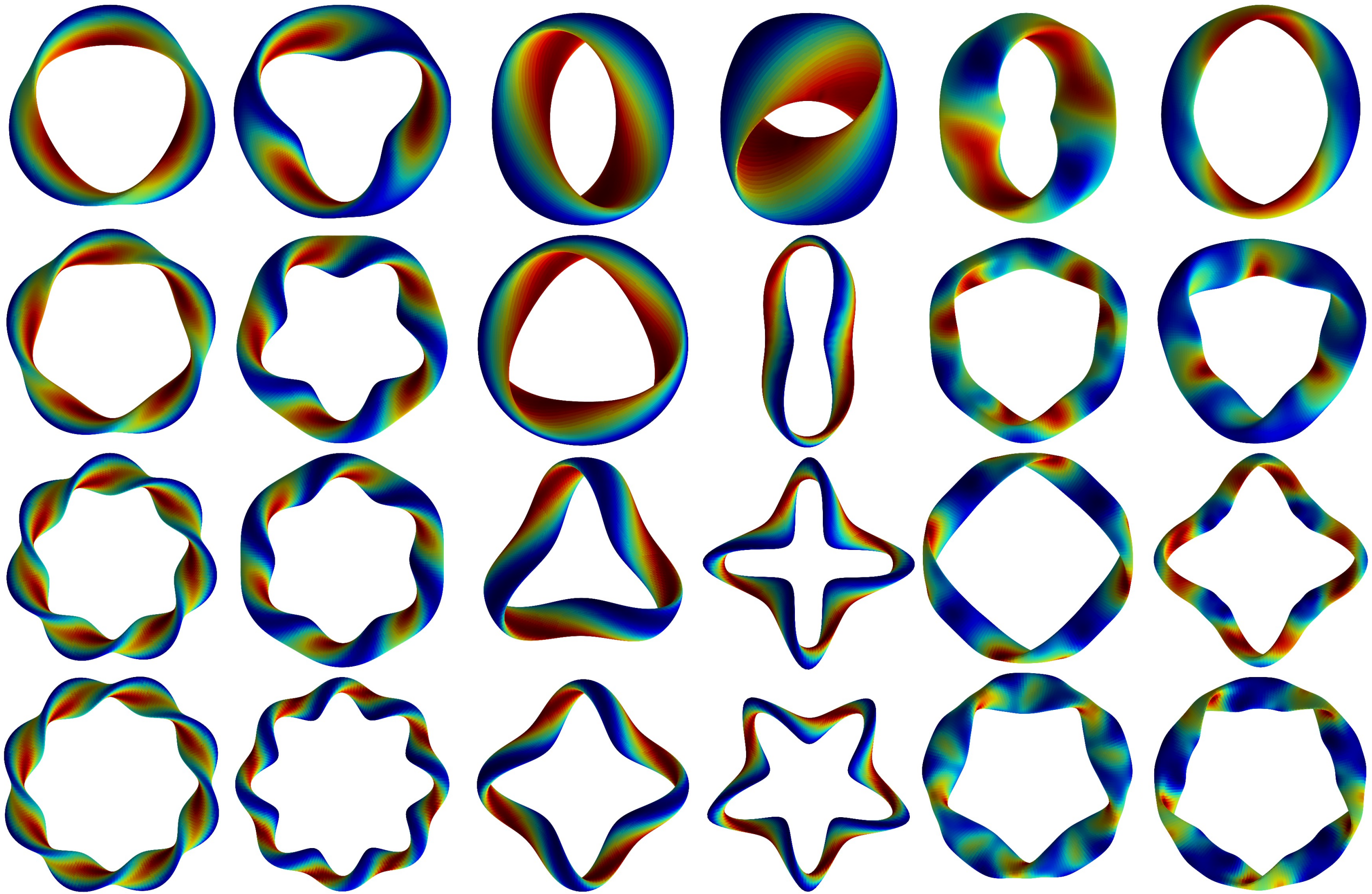
Profile prediction



Mandell (2024)


Turbulence simulations
Heat flux Q/Q_{GB}
time
Regression
R^2 = 0.989
Predicted heat flux
True heat flux
Feature importance
$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$
$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2|\nabla x|^8/B^8)$

Equilibria include rotating ellipses, quasi-symmetric, and random shapes



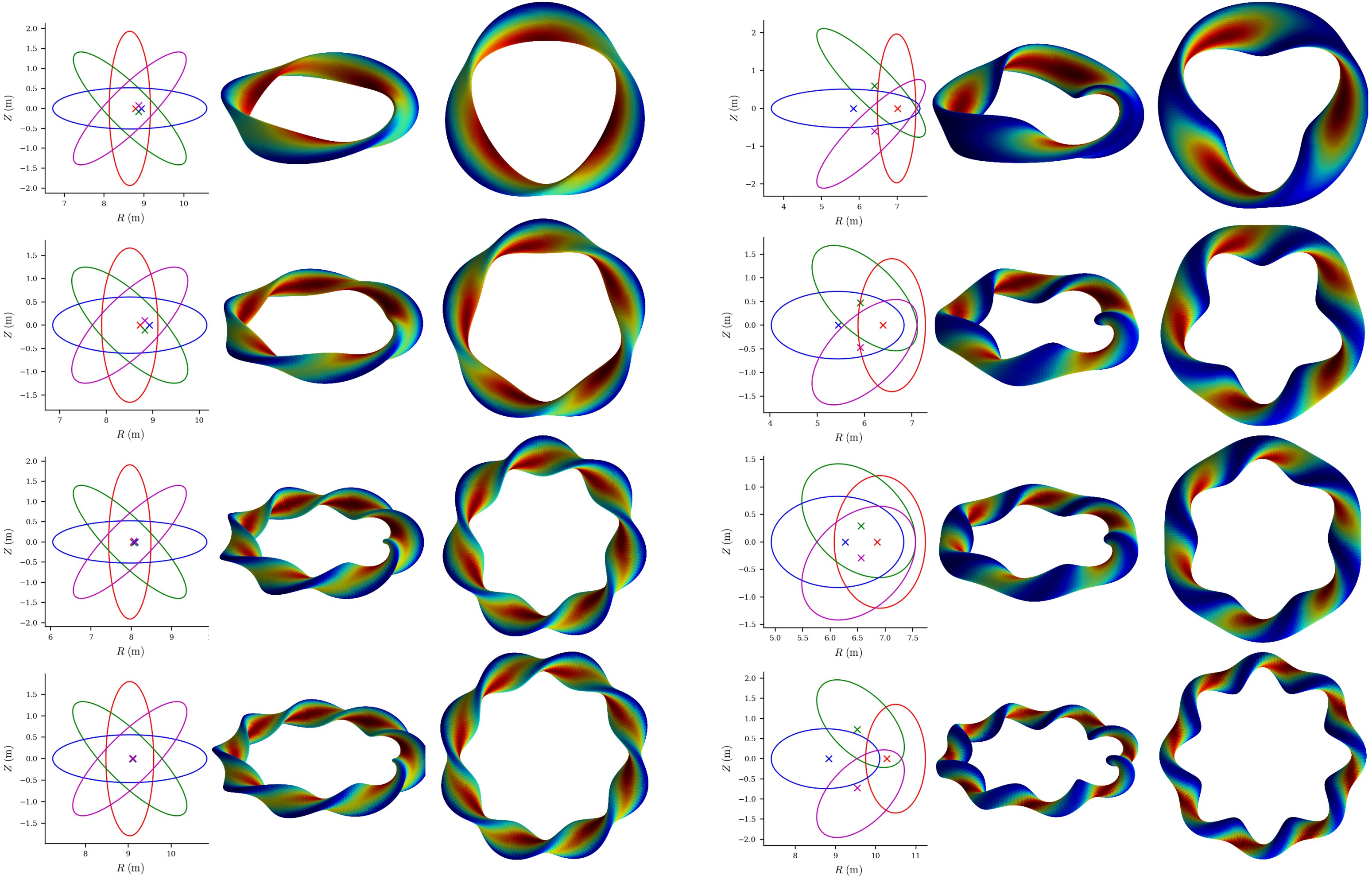
Aspect ratio, elongation, $q = 1 / \iota$, β, and number of field periods are all varied.

All configurations scaled to have same gyroBohm normalizations

(Tokamak generation underway by Ralf Mackenbach)

Turbulence simulations
Heat flux Q/Q_{GB}
time
Regression
$R^2 = 0.989$
Predicted heat flux
True heat flux
Feature importance
$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$
$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2|\nabla x|^8/B^8)$

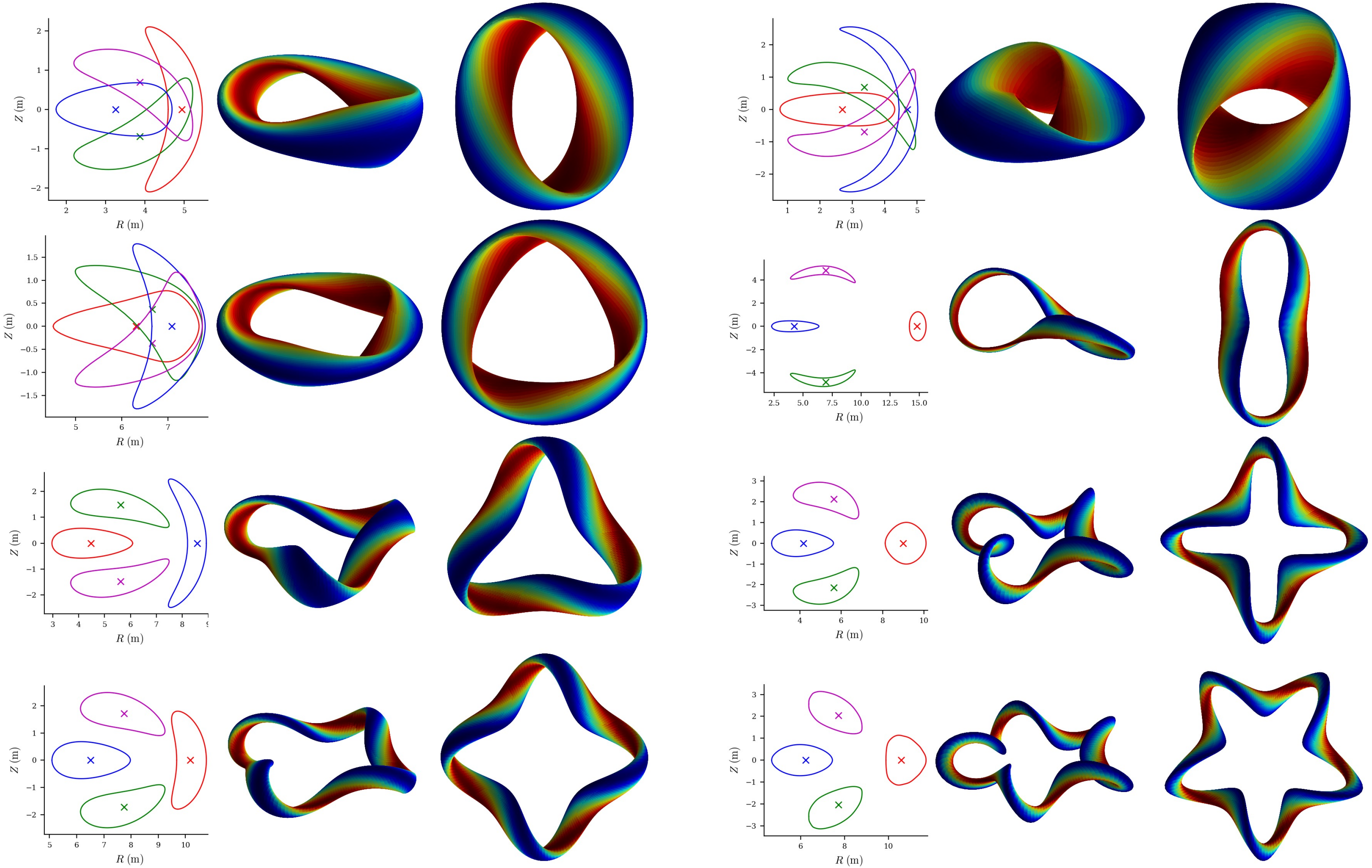
Equilibria group 1: random rotating ellipses



Nfp, aspect ratio, elongation, axis torsion, and beta are all random.

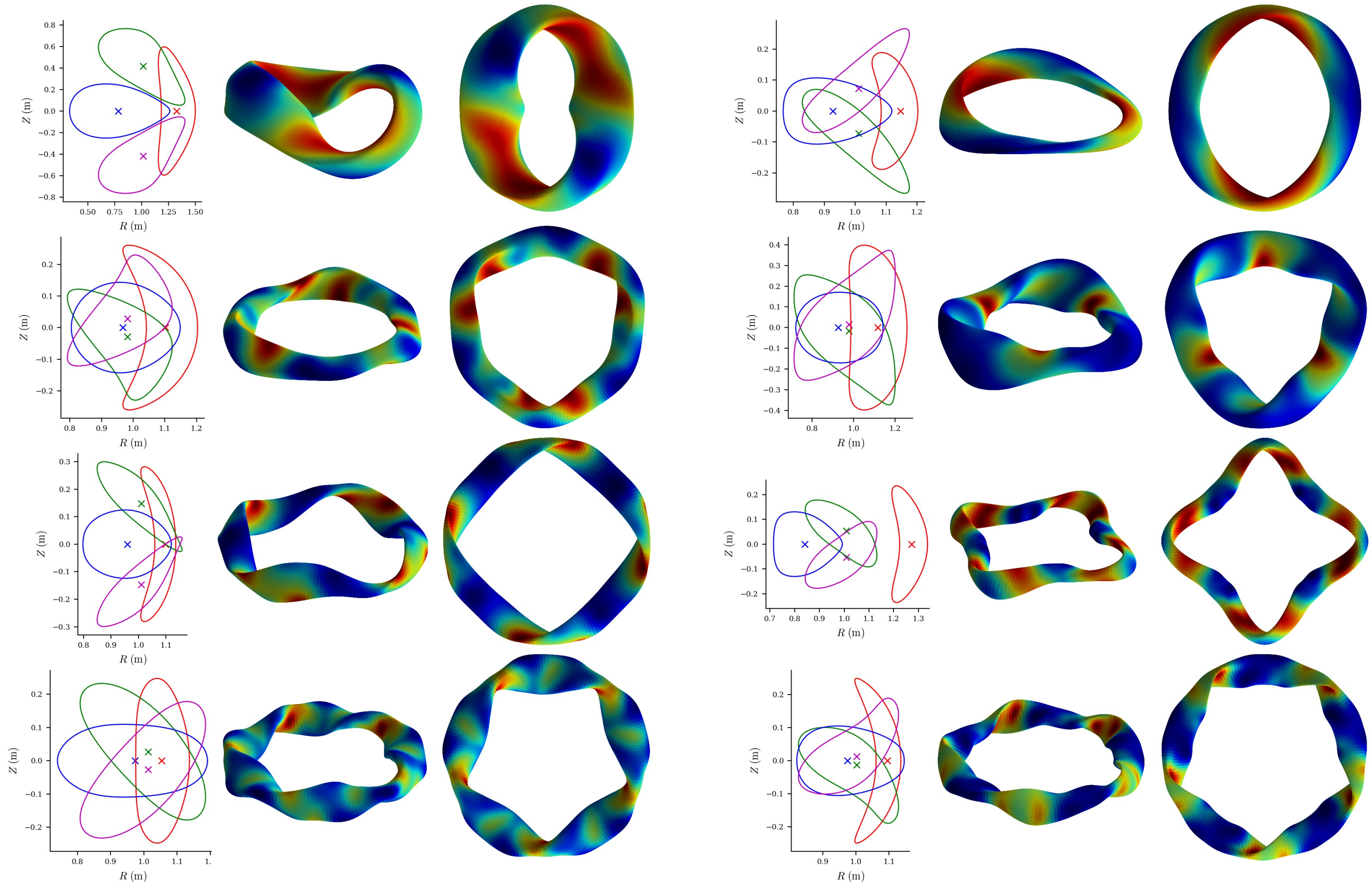
All configurations have same minor radius & toroidal flux, so same gyroBohm normalization

Equilibria group 2: QUASR QA & QH (Giuliani 2024)

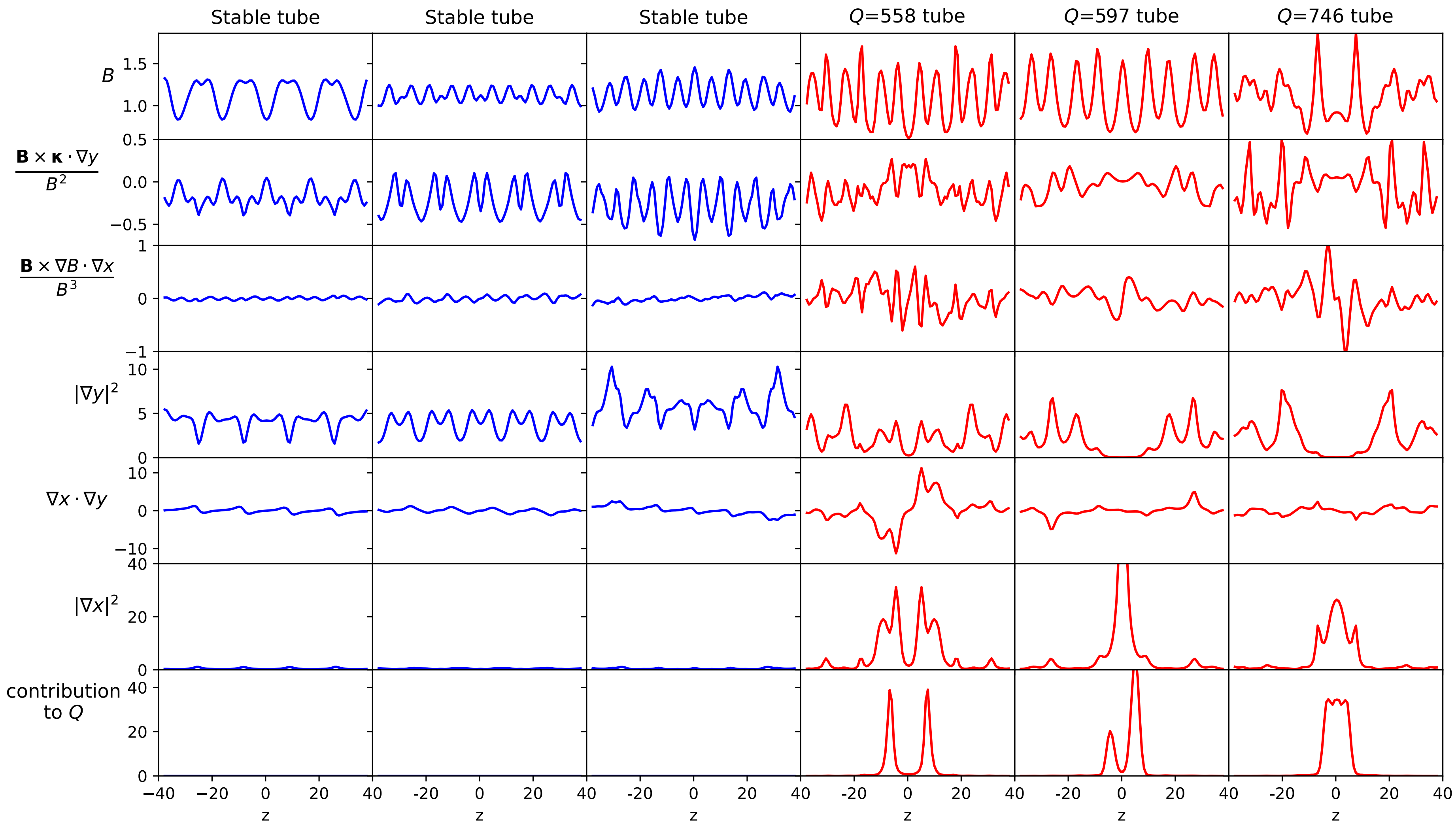


Random pressure added for even more diversity

Equilibria group 3: random boundary modes



RBC and ZBS boundary Fourier modes sampled from normal distributions, fit to 44 “real” stellarators


Stable tube
Stable tube
Stable tube
Q=558 tube
Q=597 tube
Q=746 tube
B
B × κ · ∇y / B²
B × ∇B · ∇x / B³
|∇y|²
∇x · ∇y
|∇x|²
contribution to Q
z

Our interpretable models use a large library of candidate features, all translation-invariant

Start with inputs to the gyrokinetic equation & local shear:
$F = \{B,\ B^{-3}\mathbf{B}\times\nabla B\cdot\nabla y,\ B^{-2}\mathbf{B}\times\boldsymbol{\kappa}\cdot\nabla y,\ B^{-3}\mathbf{B}\times\nabla B\cdot\nabla x,\ |\nabla x|^2,\ \nabla x\cdot\nabla y,\ |\nabla y|^2,\ d/dz(\nabla x\cdot\nabla y\ /\ |\nabla x|^2)\}$.

U = unary operations on f(z): identity, df/dz, Heaviside(f), Heaviside(-f), ReLU(f), ReLU(-f), 1/f, f^2, f/B (Jacobian), f*B

C(U(F)) = U(F) and all pairwise products of functions in U(F)

Reductions: R = {min, max, max-min, mean, median, mean square, variance, skewness, L_1 norm, quantiles 0.1, 0.25, 0.75, or 0.9, abs of fft coefficients 1-3, $k_{||}$ with largest amplitude, expected $k_{||}$, count above [-2, -1, 0, 1, 2]}

Features: R(U(C(U(F)))) $\Longrightarrow$ > 1 million combinations

Spearman correlation is a quick tool to find the most important feature

- Spearman correlation is the regular Pearson correlation of the the sorted rank of the target with the sorted rank of the feature.
- Its magnitude is invariant to any monotonic nonlinear function, e.g. corr(x, exp(x)) = 1
- No regression model required.
- Features with highest correlation to heat flux *Q* at fixed dT/dx & dn/dx:

Feature	Correlation
$\mathrm{variance}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^2/B)$	0.775
$\mathrm{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^8/B^2)$	0.774
$\mathrm{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B)$	0.772
$\mathrm{variance}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B)$	0.769
$\mathrm{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B^2)$	0.769

Heaviside function: Where there is bad curvature,

Jacobian (maybe squared)

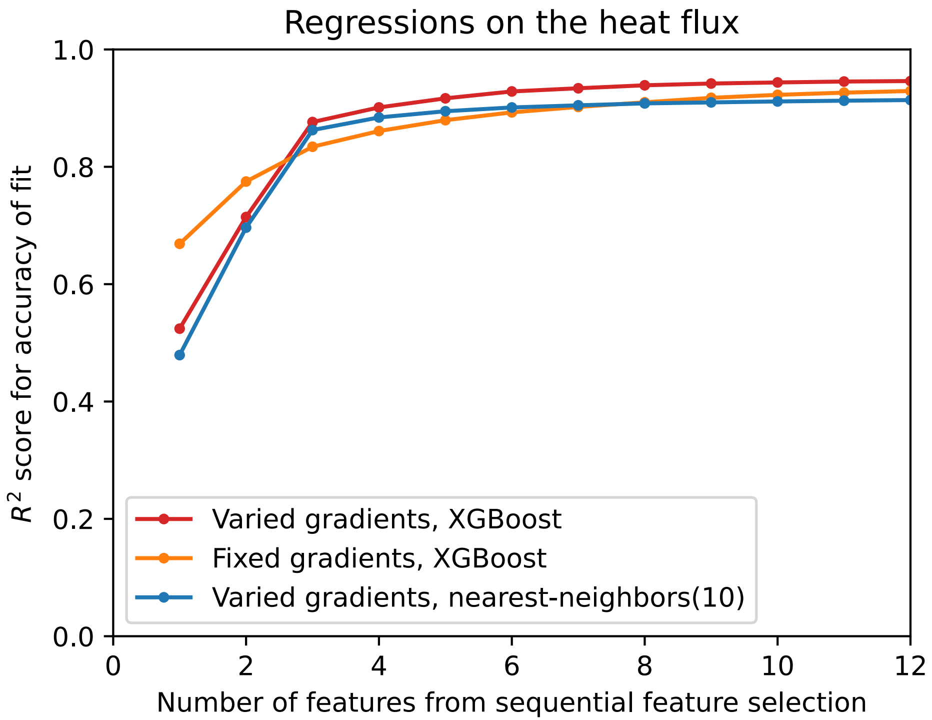
local temperature gradient in real space (to various powers)

$|\nabla T| = (dT/dx)\ |\nabla x|$

Extremely similar to Mynick (2010), Xanthopoulos (2014), Stroteich (2022), Goodman (2024)!

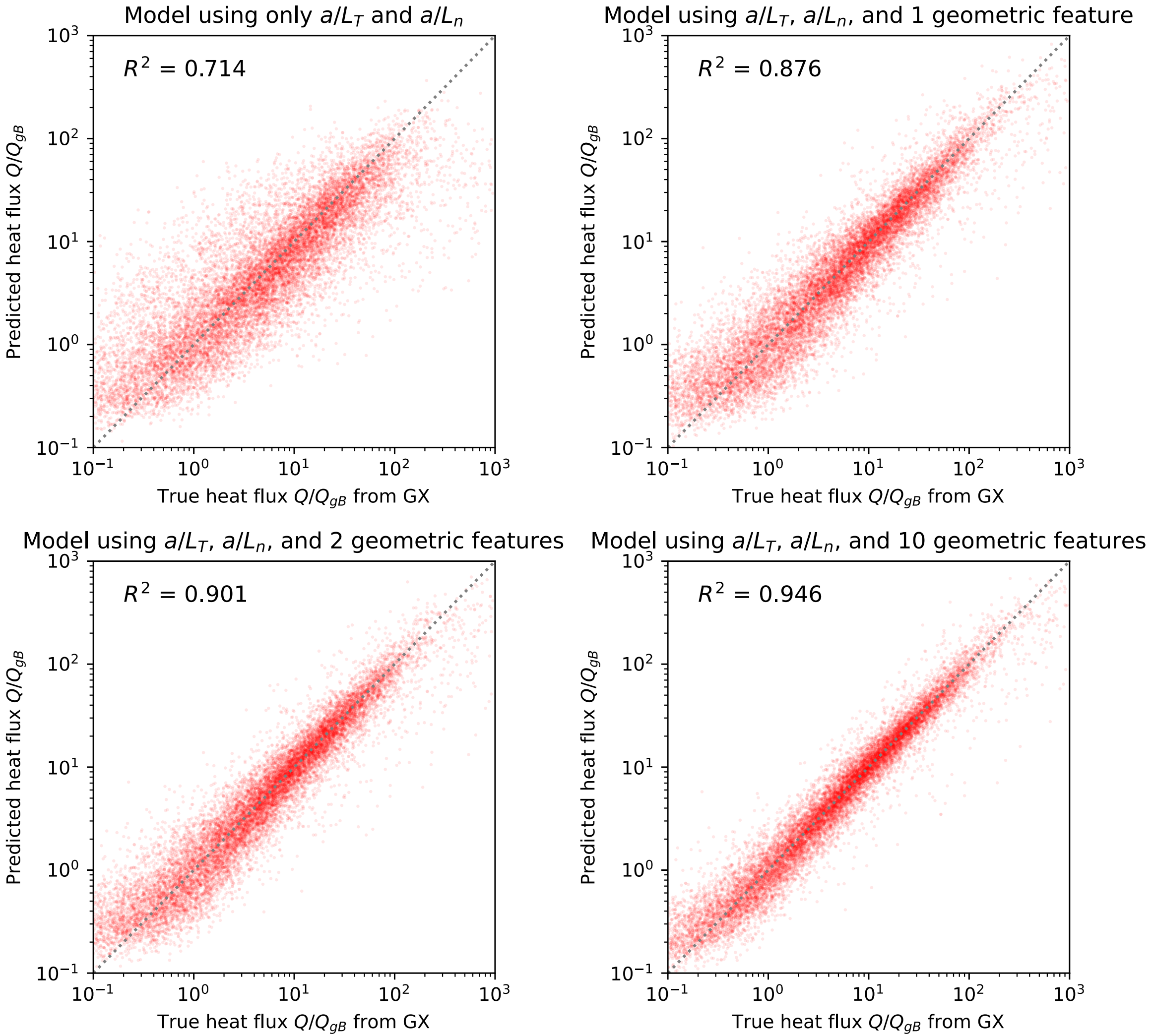
Forward sequential feature selection: ~3 features can be almost as predictive as all features

Stiffness



Critical gradient

Sequential feature selection allows closer fit to the data as more geometric features are included



Performance shown on 20% held-out test data

At each step, the top features are variations on a theme

Sequential feature selection step 3

Feature	R^2
$\mathrm{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B^2)$	0.876
$\mathrm{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B)$	0.874
$\mathrm{variance}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^2/B)$	0.871
$\mathrm{quantile0.9}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^2/B)$	0.870
$\mathrm{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^8/B^4)$	0.869

Sequential feature selection step 4

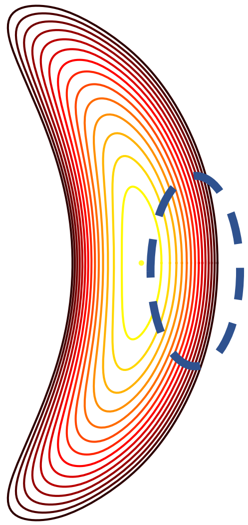
Feature	R^2
$\mathrm{median}((\mathbf{B} \times \nabla B \cdot \nabla x)^2\lvert\nabla x\rvert^8/B^8)$	0.901
$\mathrm{quantile0.75}(\mathrm{ReLU}(-\mathbf{B} \times \nabla B \cdot \nabla x)^2\lvert\nabla x\rvert^8/B^6)$	0.901
$\mathrm{median}((\mathbf{B} \times \nabla B \cdot \nabla x)^2\lvert\nabla x\rvert^8/B^6)$	0.901
$\mathrm{median}(\lvert\mathbf{B} \times \nabla B \cdot \nabla x\rvert\lvert\nabla x\rvert^4/B^4)$	0.901
$\mathrm{quantile0.75}(\mathrm{ReLU}(-\mathbf{B} \times \nabla B \cdot \nabla x)\lvert\nabla x\rvert^4/B^3)$	0.901

Regression for the random-gradient dataset

Spearman correlation is a quick tool to find the most important feature

- Unlike Pearson, Spearman is invariant to any monotonic function.
- Features with highest correlation to heat flux *Q* at fixed dT/dx & dn/dx:

Feature	Correlation
$\text{variance}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^2/B)$	0.775
$\text{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^8/B^2)$	0.774
$\text{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B)$	0.772
$\text{variance}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B)$	0.769
$\text{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^4/B^2)$	0.769



Heaviside function: Where there is bad curvature,

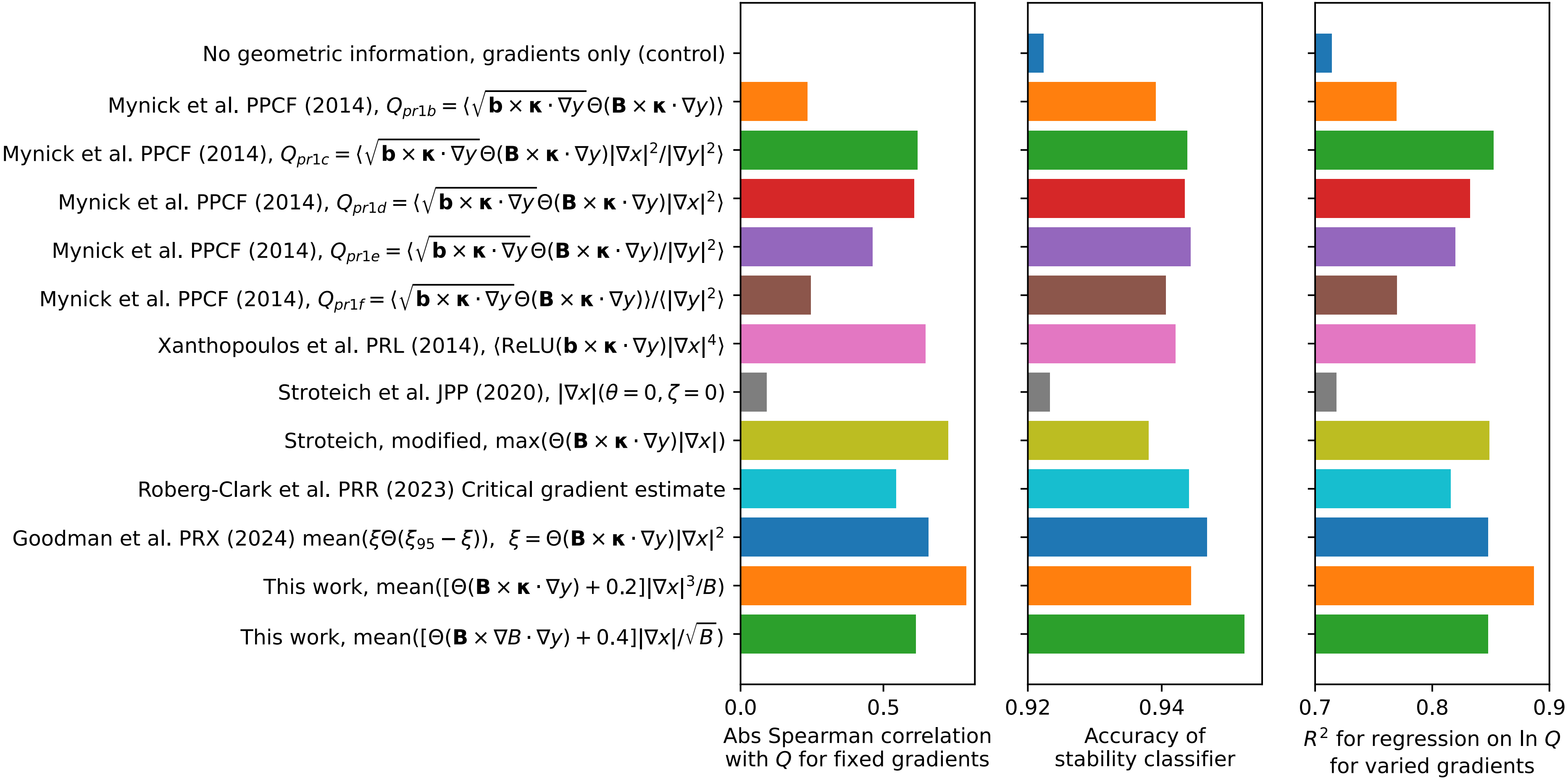
Jacobian (maybe squared)

local temperature gradient in real space (to various powers)

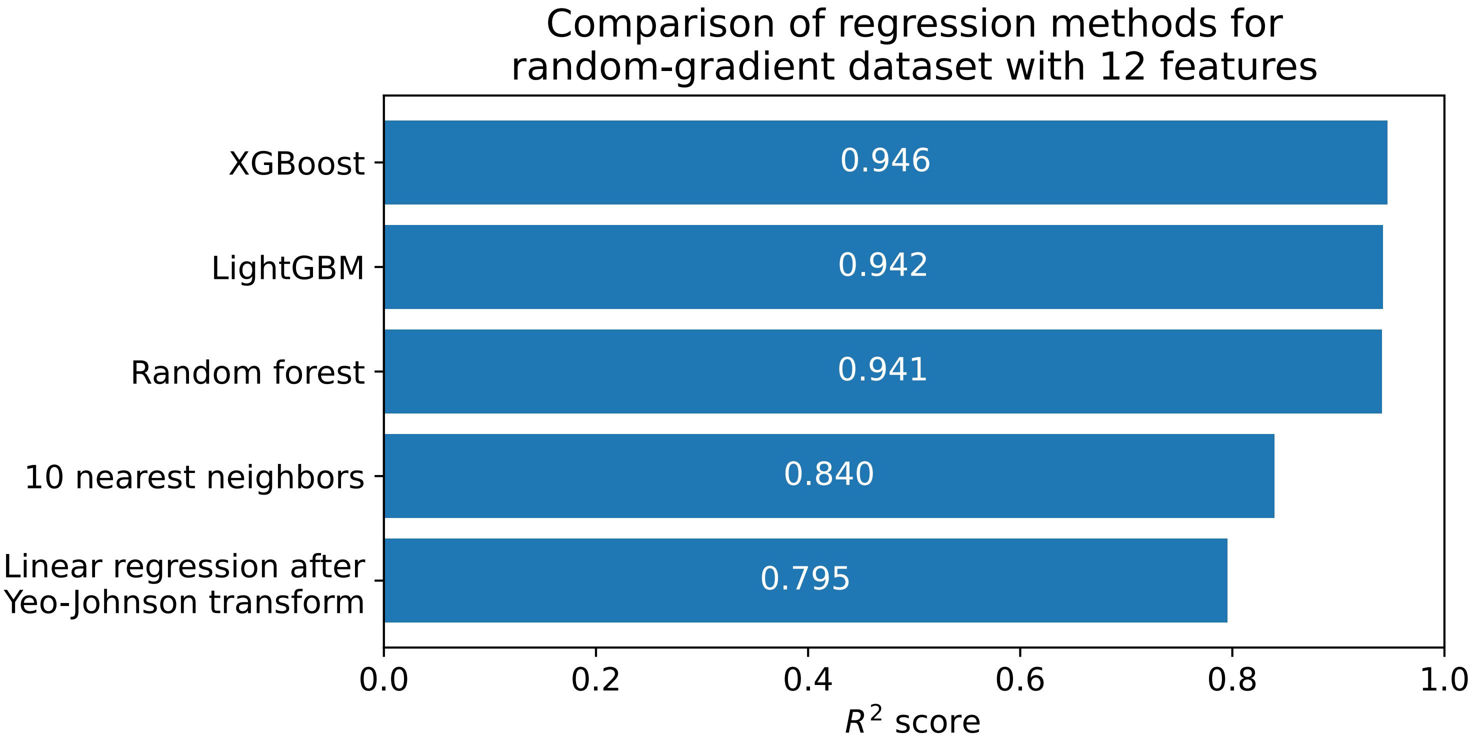
|∇T| = (dT/dx) |∇x|

Extremely similar to Mynick (2010), Xanthopoulos (2014), Stroteich (2022), Goodman (2024)!

Previously proposed proxies can be tested

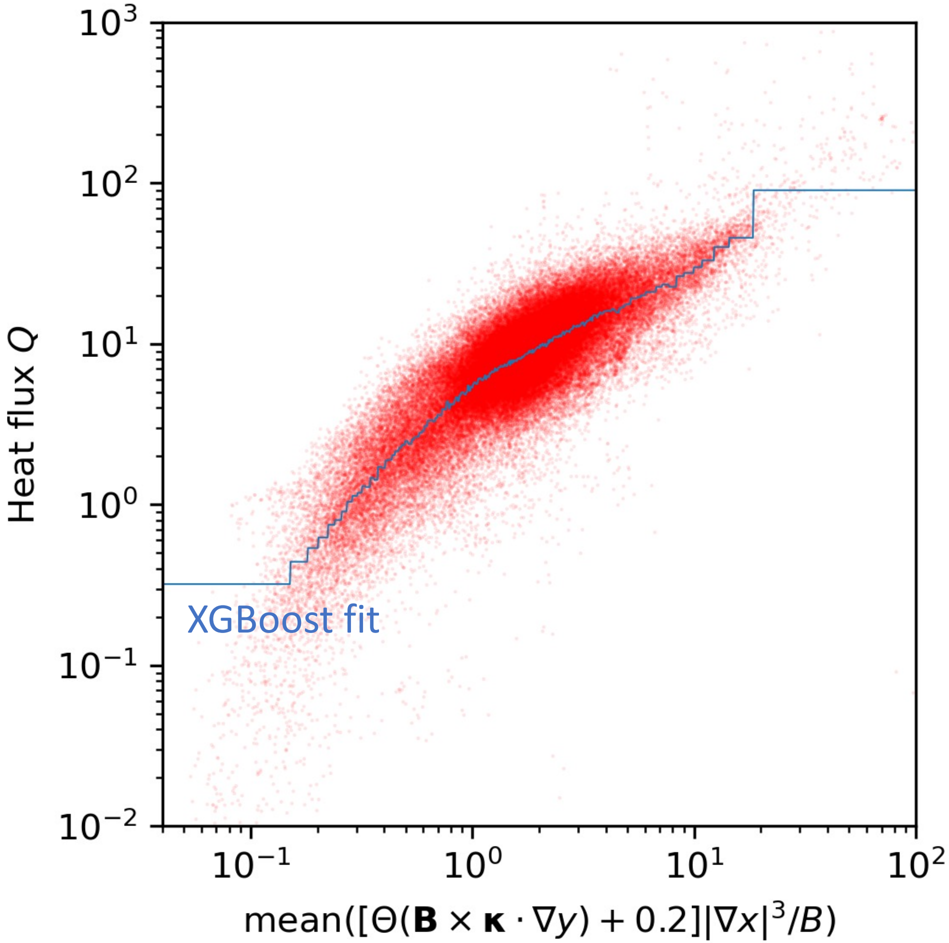


Other machine learning regression methods work also



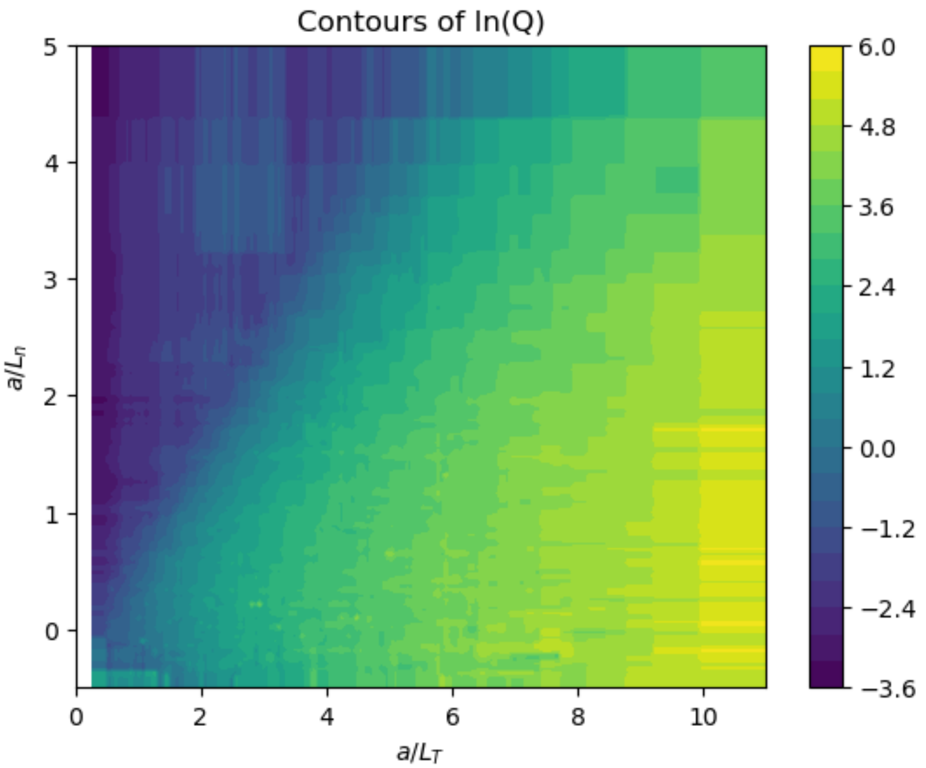
All using a/L_T, a/L_n, and the top 10 geometric features selected via XGBoost

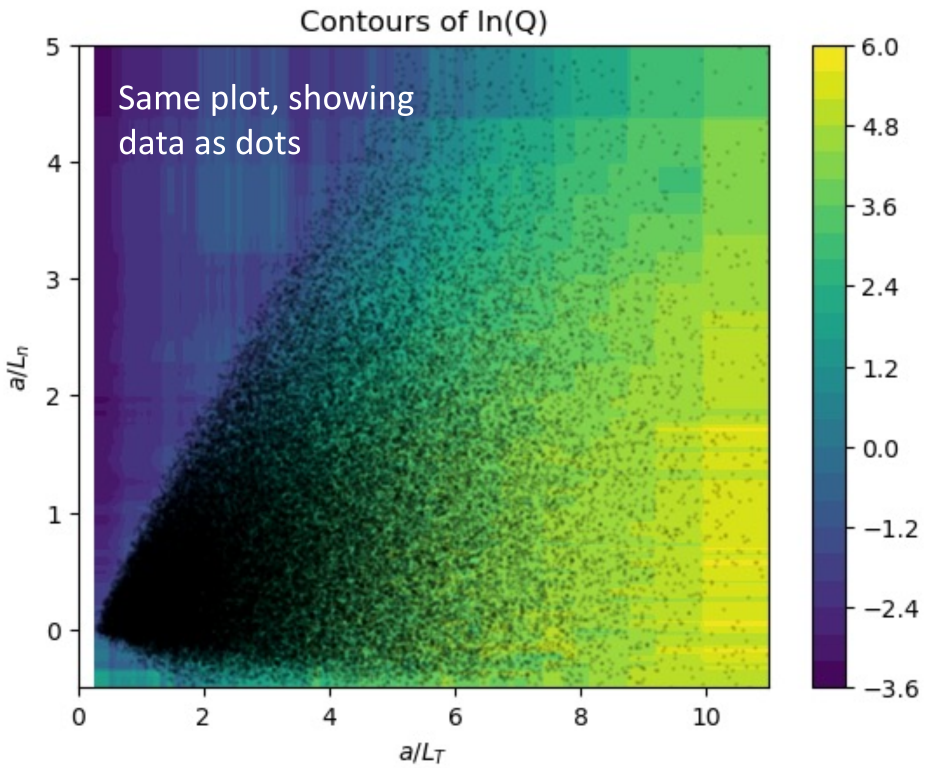
XGBoost regression model with 1 feature



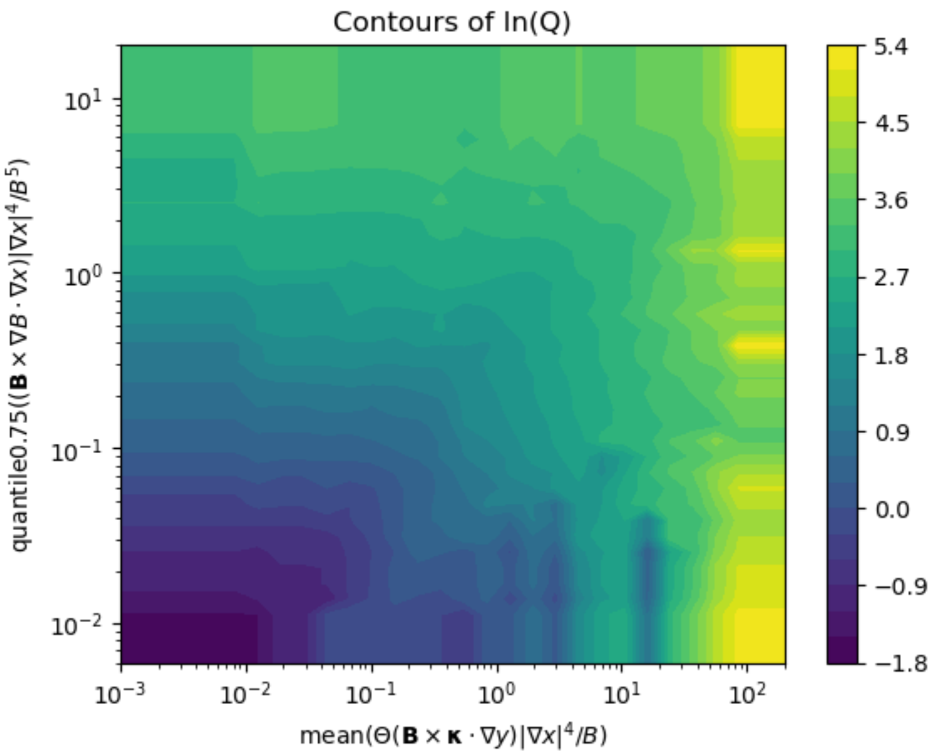
Fixed-gradient dataset

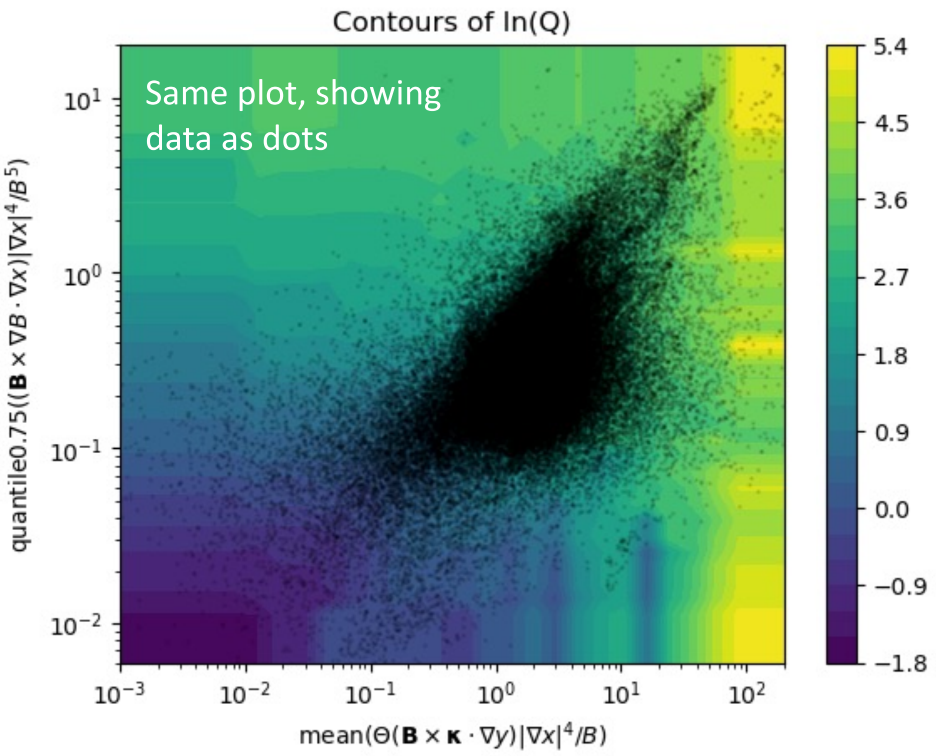
XGBoost regression model using only a/LT and a/Ln

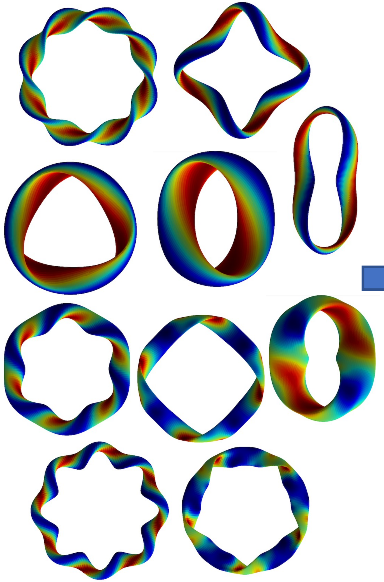


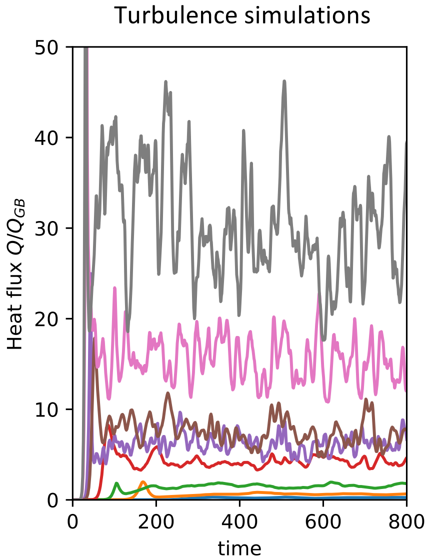


XGBoost regression model for fixed gradients using 2 features

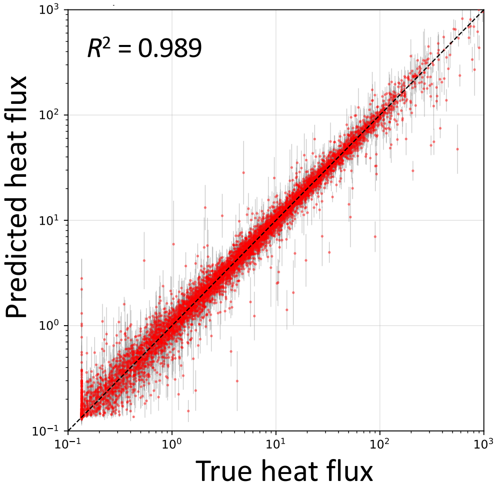








Regression

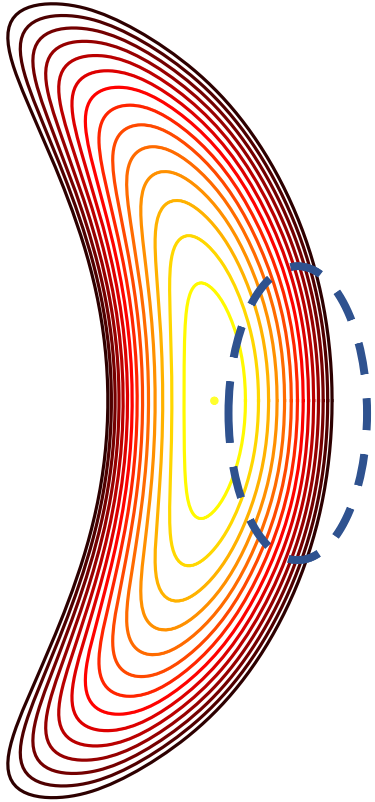


Feature importance

$$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$$
$$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2|\nabla x|^8/B^8)$$

Multiple lines of evidence agree that the most important geometric feature is $|\nabla\psi|$ in regions of bad curvature

- Highest Spearman correlation at fixed gradients.
- Consistently the first geometric feature chosen in sequential feature selection:
    - In regression on the heat flux above the critical gradient
    - And in the classifier for stability vs instability (i.e. determines critical gradient)
    - Chosen by XGBoost, nearest-neighbors, & other algorithms
- Also the largest Shapley values



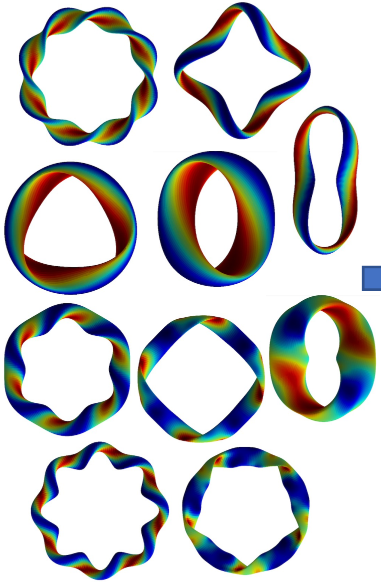
There are many extensions possible

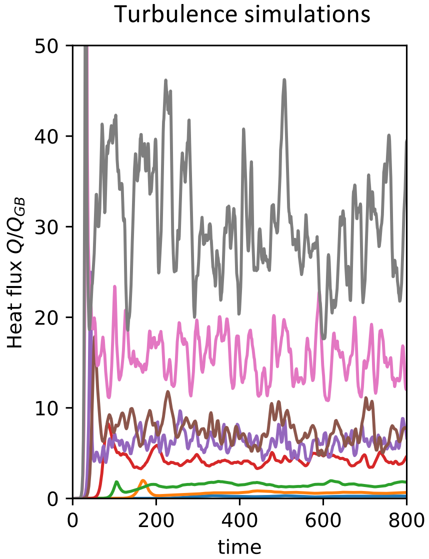
- Try larger sets of possible features
- From the gyrokinetic equation, understand how these features affect turbulence.
- Kinetic electrons, magnetic fluctuations.
- Saliency maps to understand the features learned by the neural networks.
- Symbolic regression.
- Kolmogorov-Arnold Networks.
- Optimization, profile prediction.
- Include & test other physics-motivated features.

Data is online at doi:10.5281/zenodo.14867776, so have a go at it!

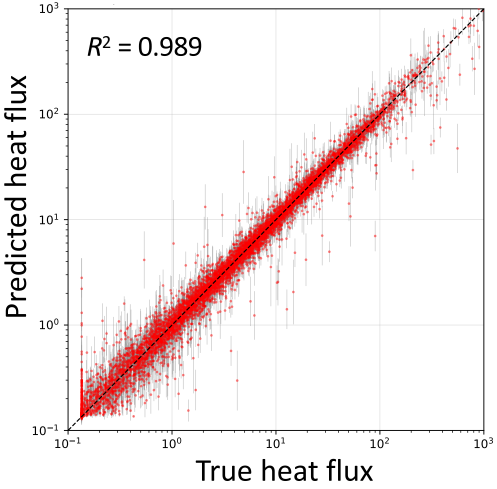
Paper: arXiv:2502.11657

Dataset doi:10.5281/zenodo.14867776





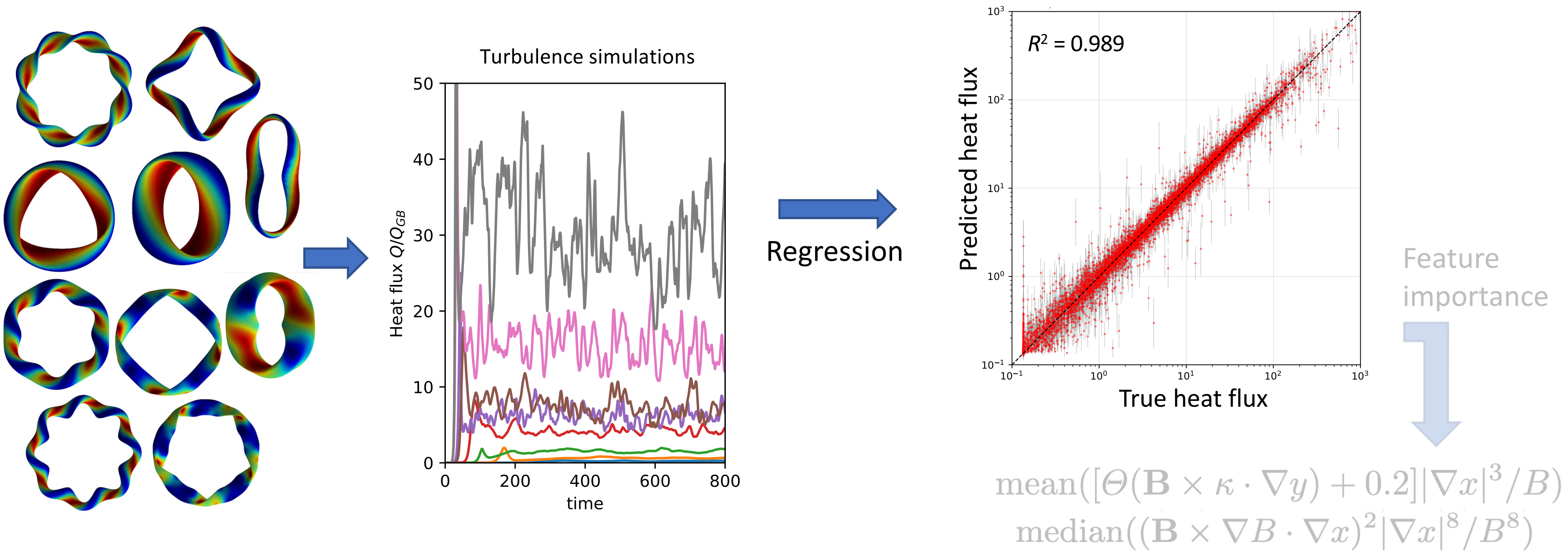
Regression



Feature importance

$$\mathrm{mean}([\Theta(\mathbf{B} \times \kappa \cdot \nabla y) + 0.2]|\nabla x|^3/B)$$

$$\mathrm{median}((\mathbf{B} \times \nabla B \cdot \nabla x)^2 |\nabla x|^8/B^8)$$


Turbulence simulations
Heat flux Q/Q_{GB}
time
Regression
R^2 = 0.989
Predicted heat flux
True heat flux
Feature importance
$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$
$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2|\nabla x|^8/B^8)$

To create library of candidate features, apply any translation-invariant reduction to translation-equivariant operations

Start with inputs to the gyrokinetic equation & local shear:

B, $B^{-3}\mathbf{B}\times\nabla B\cdot\nabla y$, $B^{-2}\mathbf{B}\times\boldsymbol{\kappa}\cdot\nabla y$, $B^{-3}\mathbf{B}\times\nabla B\cdot\nabla x$,
$|\nabla x|^2$, $\nabla x\cdot\nabla y$, $|\nabla y|^2$, $d/dz(\nabla x\cdot\nabla y \,/\, |\nabla x|^2)$.

Apply unary operations on $f(z)$: f^2, df/dz, Heaviside(f), etc.,

Include all pairwise products,

Reductions: max, median, abs of fft coefficients, $k_{||}$ with largest amplitude, etc.

$\Longrightarrow$ > 1 million combinations

Most important features from sequential feature selection

Regression on heat flux

Feature	R^2
a/L_T	0.524
a/L_n	0.714
$\mathrm{mean}(\Theta(\mathbf{B}\times\kappa\cdot\nabla y)\lvert\nabla x\rvert^4/B^2)$	0.876
$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2\lvert\nabla x\rvert^8/B^8)$	0.901
$\mathrm{absFFTCoeff1}(\mathrm{ReLU}(\mathbf{B}\times\nabla B\cdot\nabla x)/B^5)$	0.917

Classification (stability vs instability)

Feature	log-loss
a/L_T	0.361
a/L_n	0.189
$\mathrm{mean}(\Theta(\mathbf{B}\times\nabla B\cdot\nabla y)\lvert\nabla x\rvert^2/B)$	0.122
$\mathrm{mean}(\Theta(-\mathbf{B}\times\nabla B\cdot\nabla x)\lvert\nabla x\rvert^2 B)$	0.105
$\mathrm{mean}((\mathbf{B}\times\kappa\cdot\nabla y)/B)$	0.094

The gradients are more important than any geometric feature

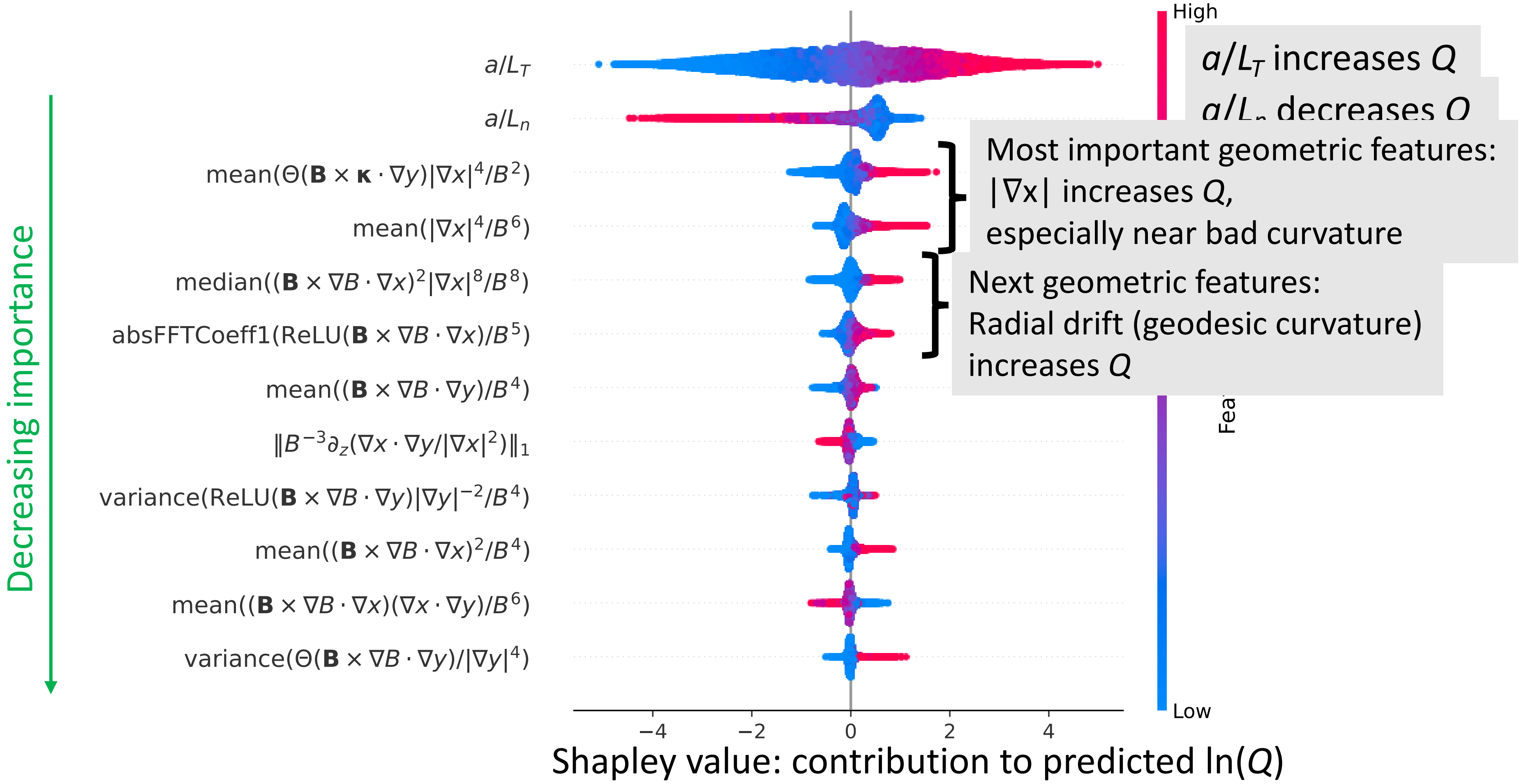
The most important geometric feature is flux surface compression where curvature is bad

The 2nd most important geometric feature involves flux surface compression and radial ∇B drift

Xanthopoulos et al (2011), Nakata & Matsuoka (2022):

Larger geodesic curvature (= radial drift) ⇒ Stronger damping of zonal flows ⇒ higher heat flux

Shapley values show the sign and magnitude of each feature's effect



Summary

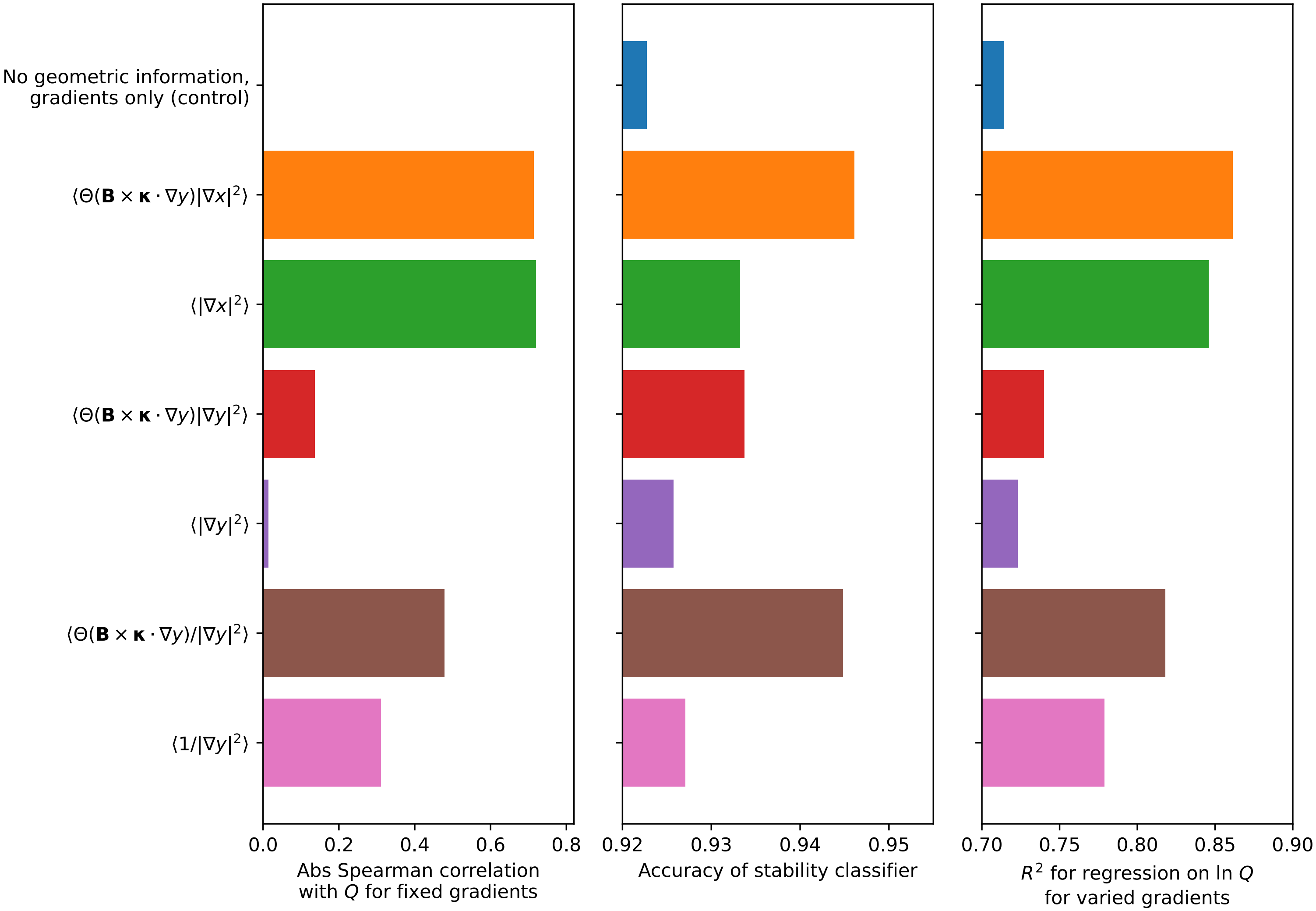
- Interpretable ML can reveal trends and stimulate theory.
- Most important feature for ITG seems to be $|\nabla\psi|$ in regions of bad curvature.

Future work

- From the gyrokinetic equation, understand how top features affect turbulence.
- Saliency maps to understand the features learned by the neural networks.
- Other interpretable methods (symbolic regression, Kolmogorov-Arnold Networks)
- Kinetic electrons, magnetic fluctuations.
- Optimization & profile prediction.

Paper:
arXiv:2502.11657

Dataset
doi:10.5281/zenodo.14867776



No geometric information, gradients only (control)
$\langle\Theta(\mathbf{B}\times\boldsymbol{\kappa}\cdot\nabla y)|\nabla x|^2\rangle$
$\langle|\nabla x|^2\rangle$
$\langle\Theta(\mathbf{B}\times\boldsymbol{\kappa}\cdot\nabla y)|\nabla y|^2\rangle$
$\langle|\nabla y|^2\rangle$
$\langle\Theta(\mathbf{B}\times\boldsymbol{\kappa}\cdot\nabla y)/|\nabla y|^2\rangle$
$\langle 1/|\nabla y|^2\rangle$
0.0 0.2 0.4 0.6 0.8
Abs Spearman correlation with Q for fixed gradients
0.92 0.93 0.94 0.95
Accuracy of stability classifier
0.70 0.75 0.80 0.85 0.90
R^2 for regression on $\ln Q$ for varied gradients