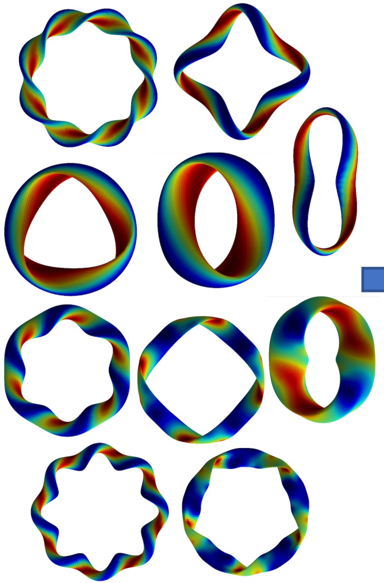
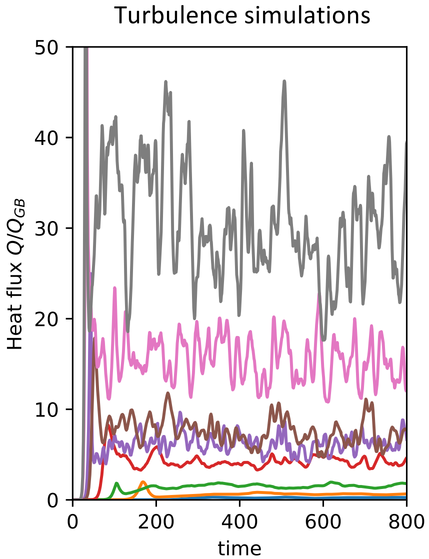
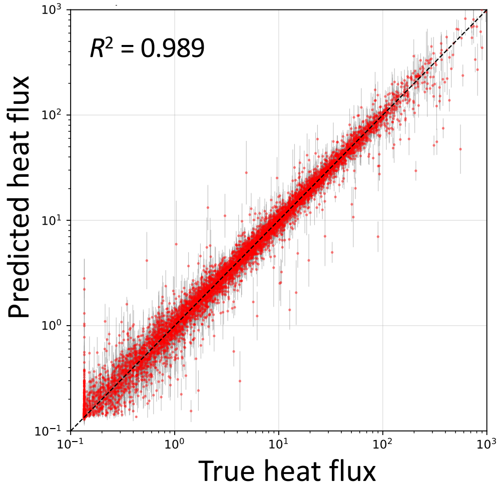
# How does magnetic geometry affect ITG turbulence? Insights from data & machine learning






Regression




Feature importance

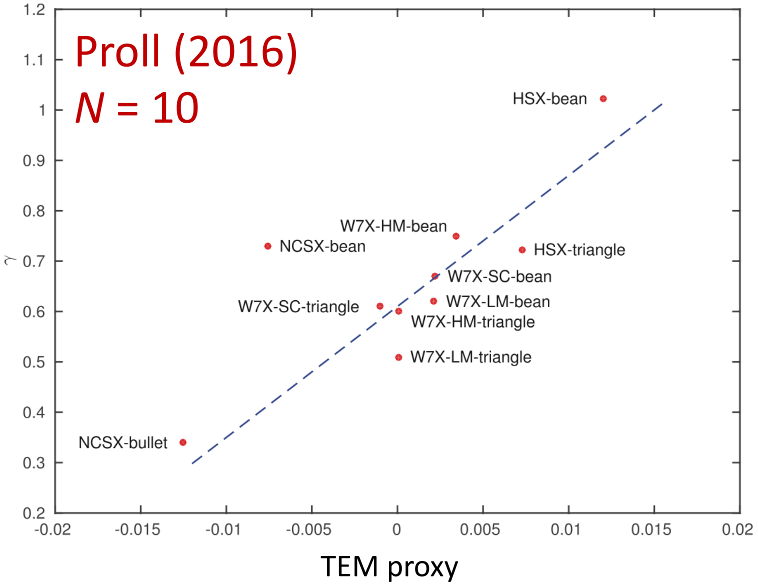
$$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$$

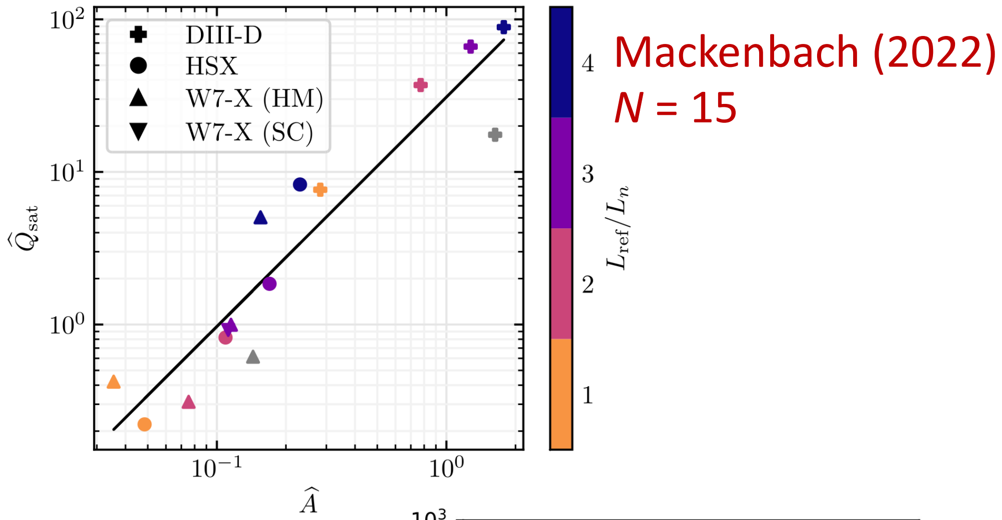
$$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2|\nabla x|^8/B^8)$$

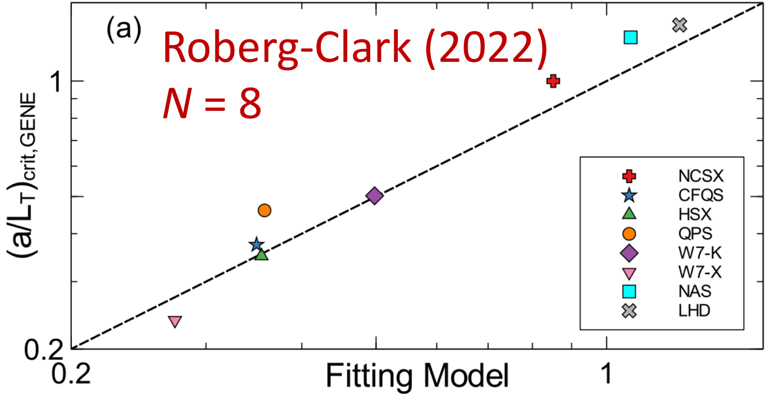
M Landreman, J Y Choi, C Alves, P Balaprakash, R M Churchill , R Conlin, G Roberg-Clark

Thanks to many others who gave suggestions

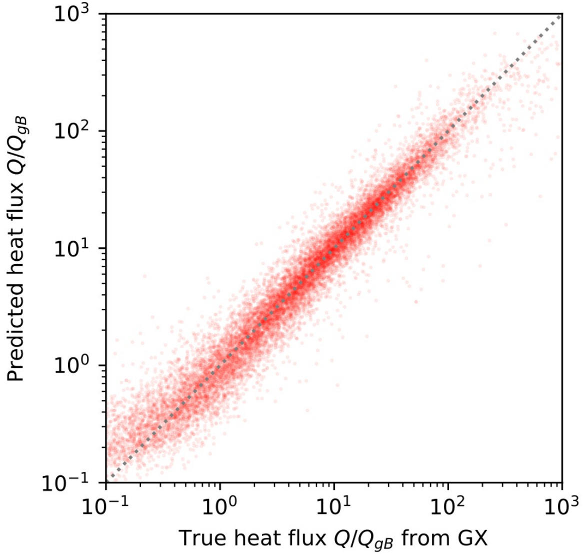
*Supported by the US DOE StellFoundry SciDAC*



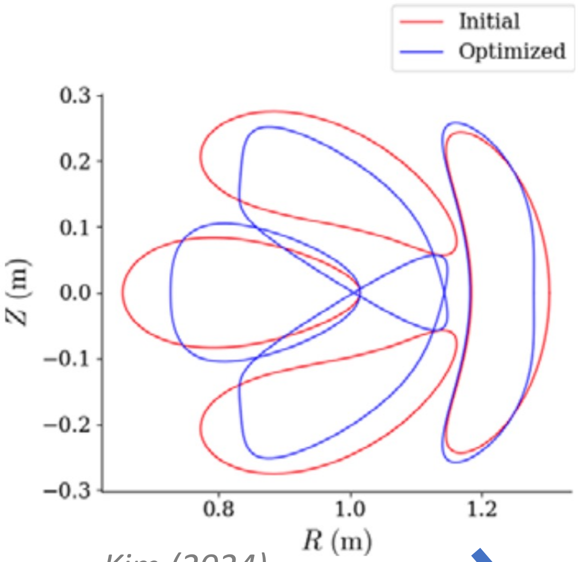

This work:
*N* = 100,705



# Motivations

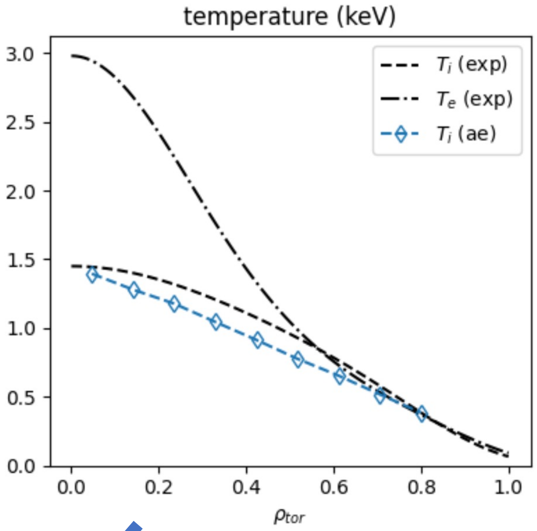
Understanding

Optimization




Kim (2024)

Profile prediction

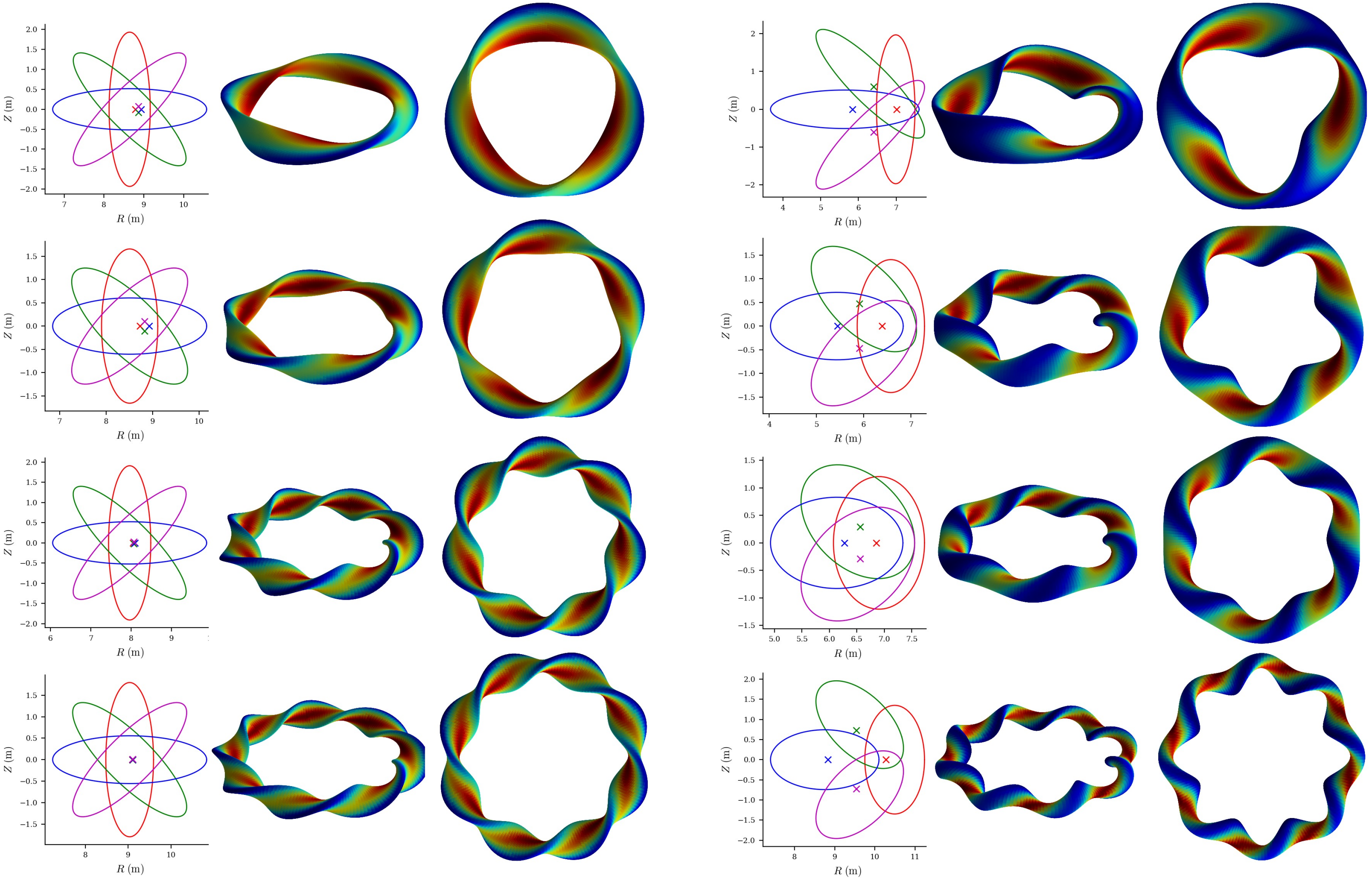



Mandell (2024)

Optimize geometry for maximum fusion power

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

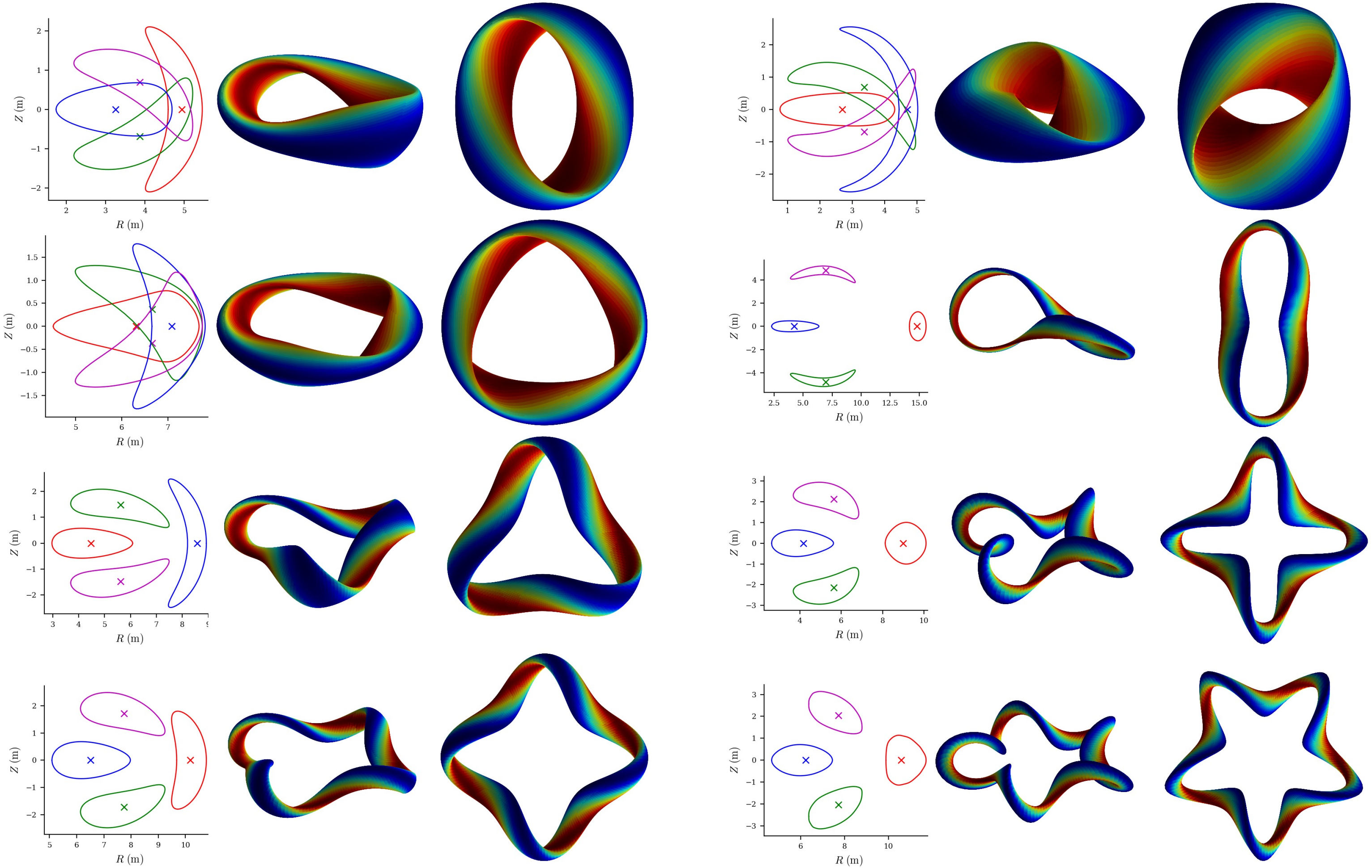
# Equilibria group 1: random rotating ellipses




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

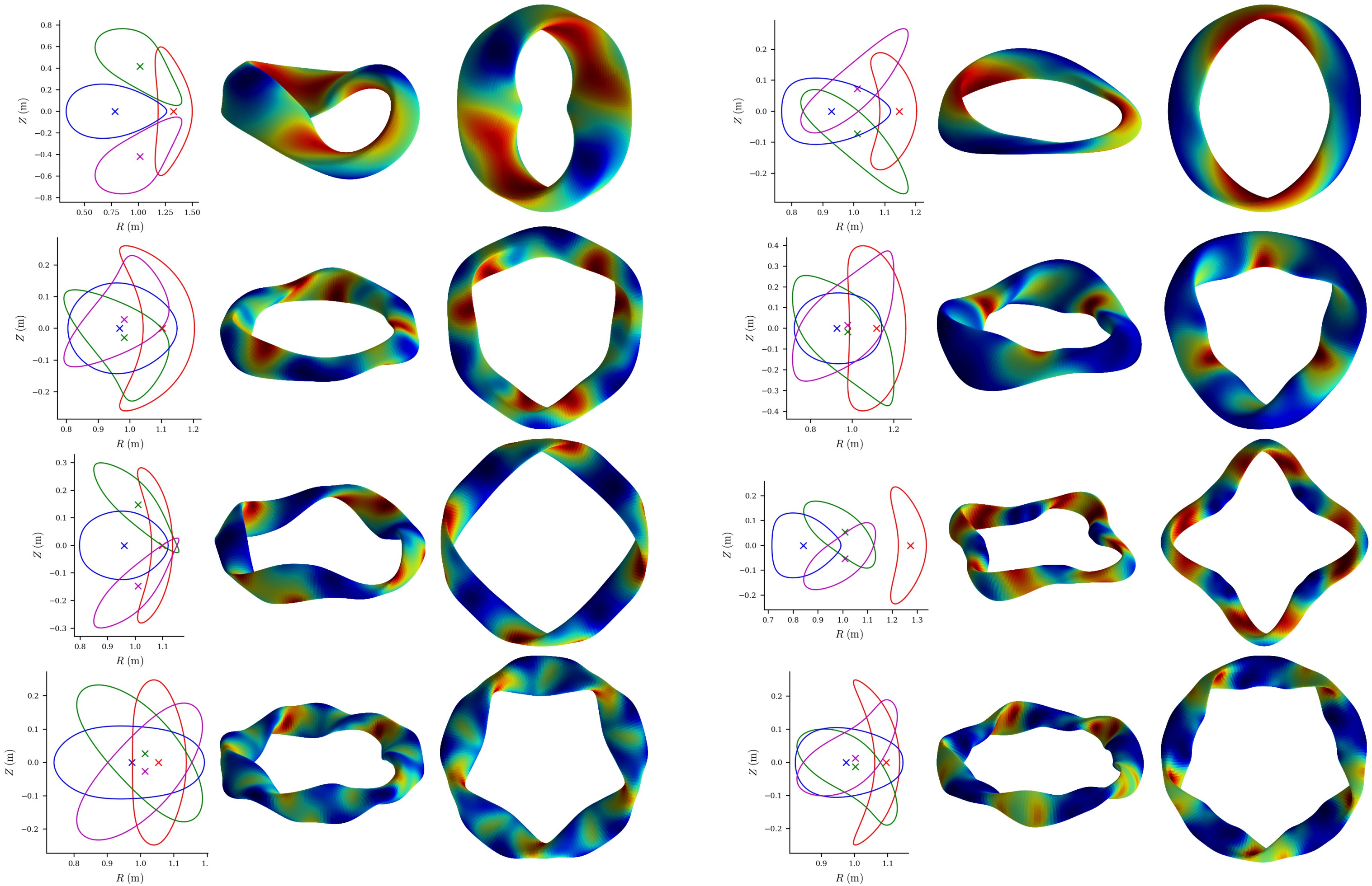
All configurations have same minor radius & toroidal flux, so same gyroBohm normalization

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

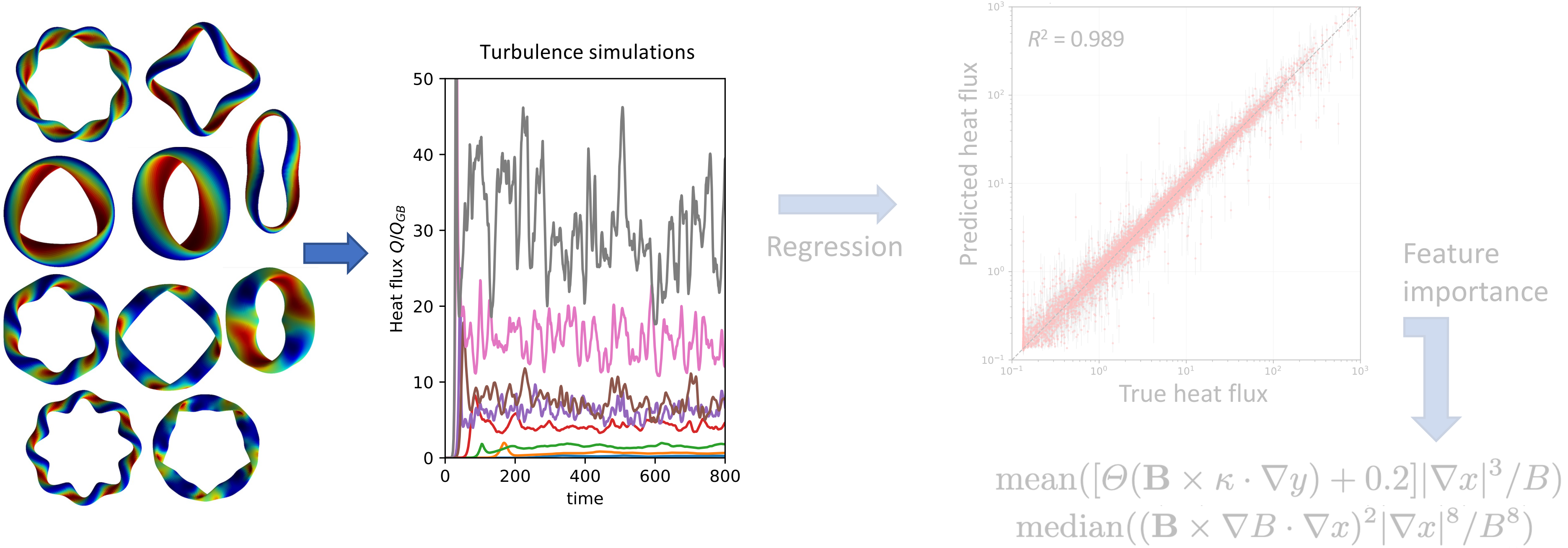



Random pressure added for even more diversity

# Equilibria group 3: random boundary modes

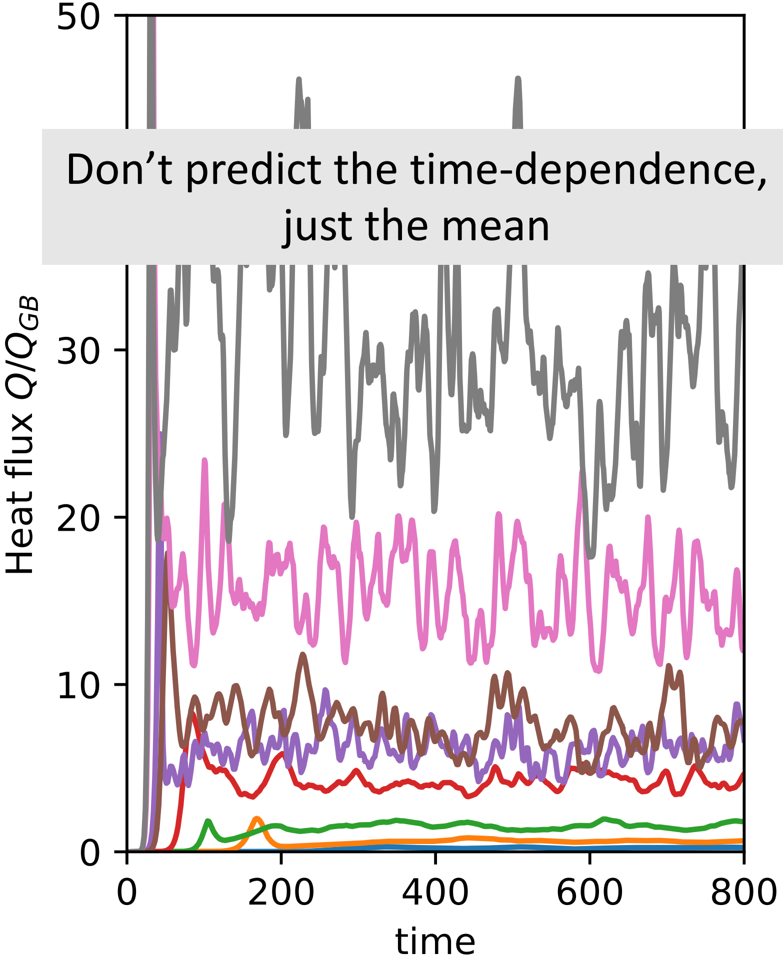



RBC and ZBS boundary Fourier modes sampled from normal distributions, fit to 44 “real” stellarators

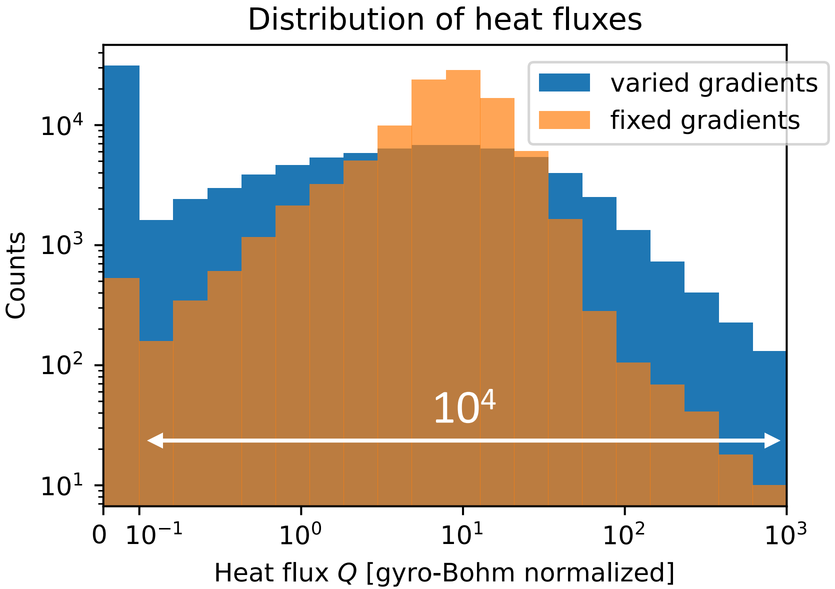


Turbulence simulations
Heat flux $Q/Q_{GB}$
time
Regression
$R^2 = 0.989$
Predicted heat flux
True heat flux
Feature importance
$\mathrm{mean}([\Theta(\mathbf{B} \times \kappa \cdot \nabla y) + 0.2]|\nabla x|^3/B)$
$\mathrm{median}((\mathbf{B} \times \nabla B \cdot \nabla x)^2|\nabla x|^8/B^8)$

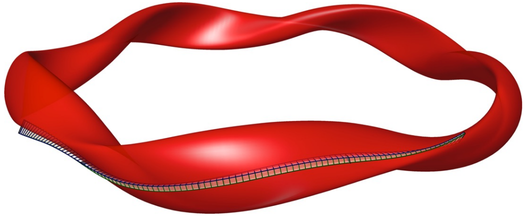
# Nonlinear turbulence simulations were run with GX in every equilibrium

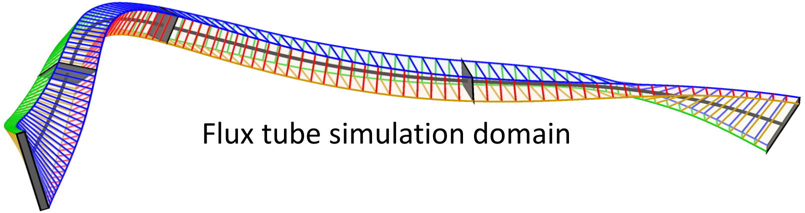



- Electrostatic, adiabatic electrons.
- 1 simulation in each tube with random dT/dx and dn/dx.
- 1 simulation in each tube with (a/T) dT/dx = 3, (a/n) dn/dx=0.9
- 8 minutes to get heat flux on 1 GPU
- $2\times10^5$ nonlinear simulations took < 7000 node-hours (1/8 allocation)



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

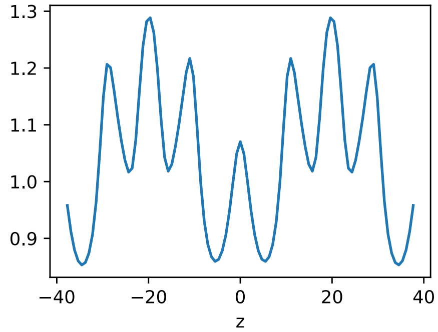




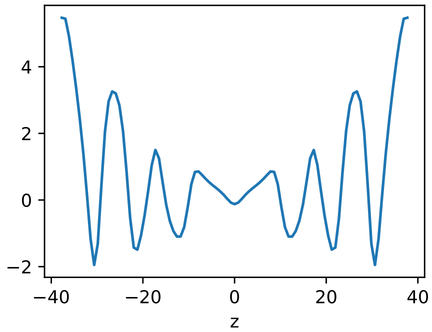

$$\boldsymbol{B} = B_{ref} \nabla x \times \nabla y$$

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

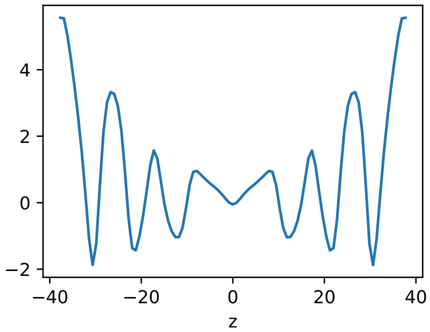
$|B|$



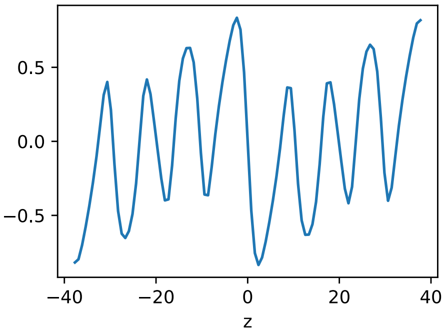

$|B|^{-3}\boldsymbol{B}\times\nabla|B|\cdot\nabla y$



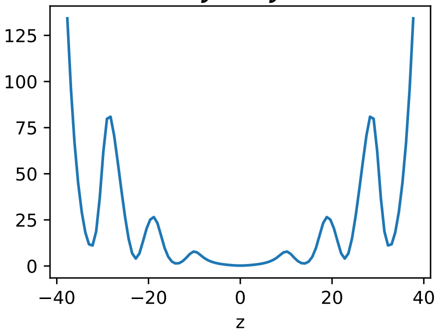

$|B|^{-2}\boldsymbol{B}\times\boldsymbol{\kappa}\cdot\nabla y$



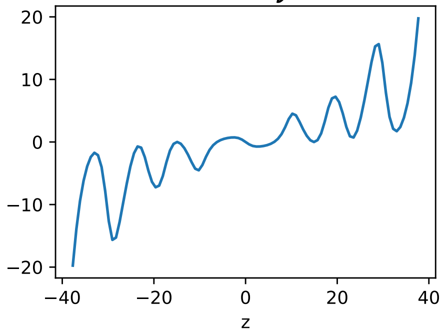

$|B|^{-3}\boldsymbol{B}\times\nabla|B|\cdot\nabla x$



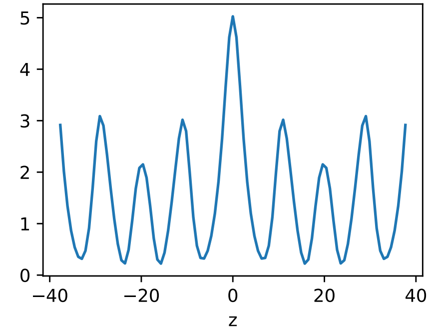

$\nabla y \cdot \nabla y$



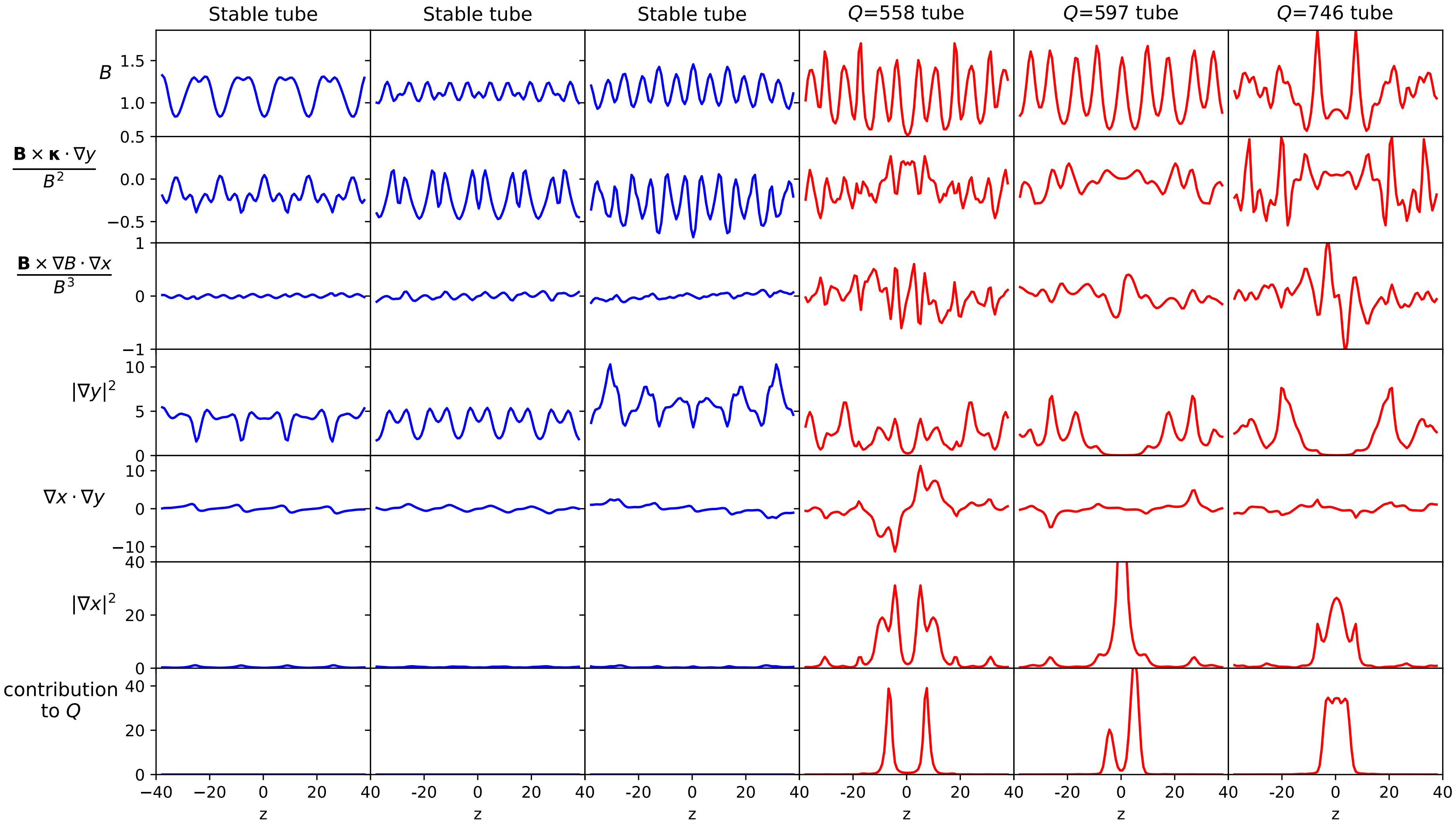

$\nabla x \cdot \nabla y$



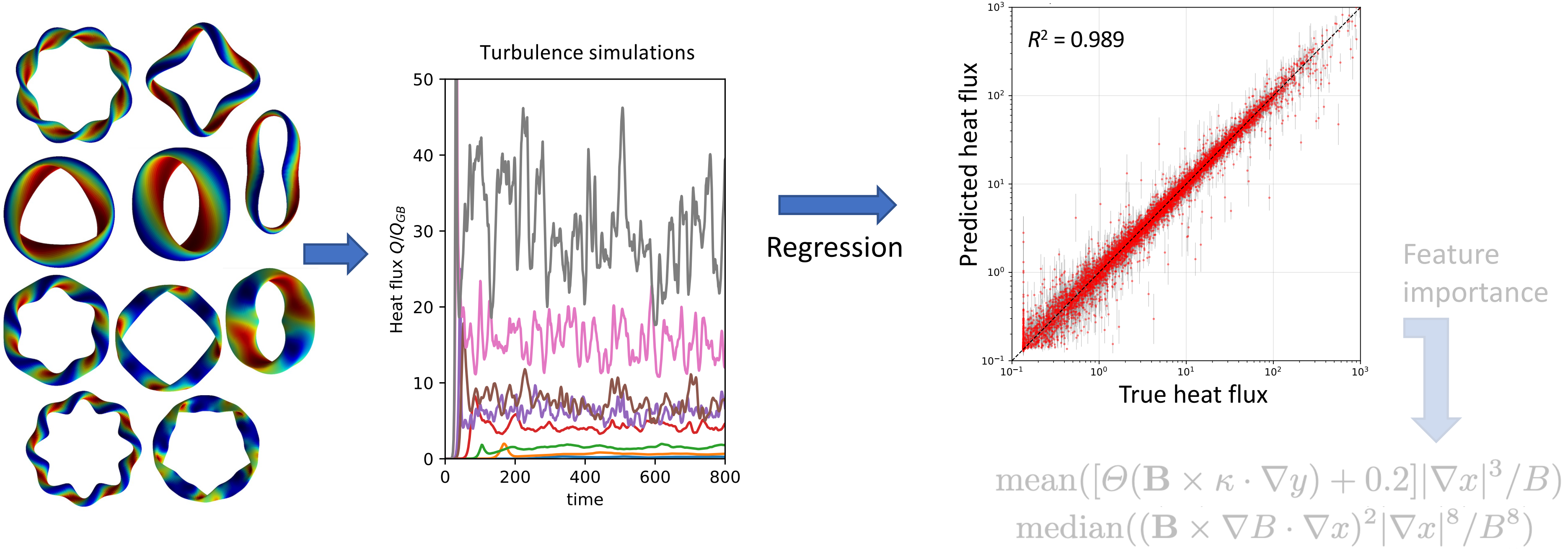

$\nabla x \cdot \nabla x$




$\boldsymbol{b}\cdot\nabla z$ is constant and the same for all configs, as are tube lengths in meters, so Fourier modes ($k_{||}$) can be compared between configurations.



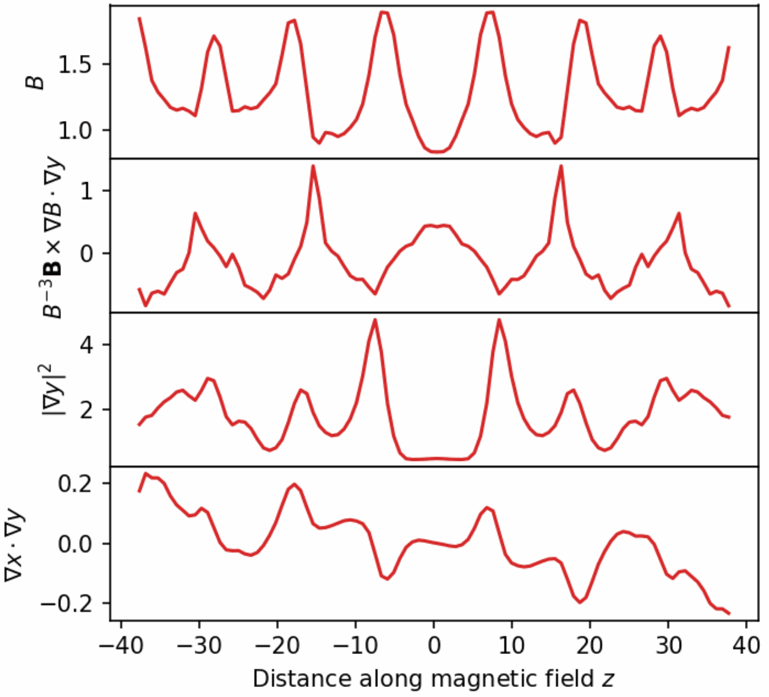
Stable tube
Stable tube
Stable tube
Q=558 tube
Q=597 tube
Q=746 tube
B
B × κ · ∇y / B^2
B × ∇B · ∇x / B^3
|∇y|^2
∇x · ∇y
|∇x|^2
contribution to Q
z

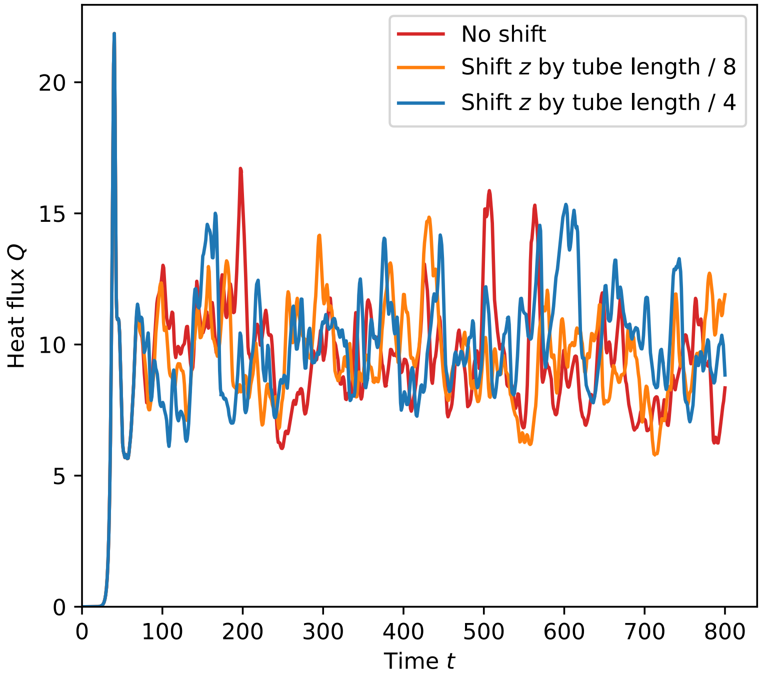


Turbulence simulations
Heat flux $Q/Q_{GB}$
time
Regression
$R^2$ = 0.989
Predicted heat flux
True heat flux
Feature importance
$\mathrm{mean}([\Theta(\mathbf{B}\times\kappa\cdot\nabla y)+0.2]|\nabla x|^3/B)$
$\mathrm{median}((\mathbf{B}\times\nabla B\cdot\nabla x)^2|\nabla x|^8/B^8)$

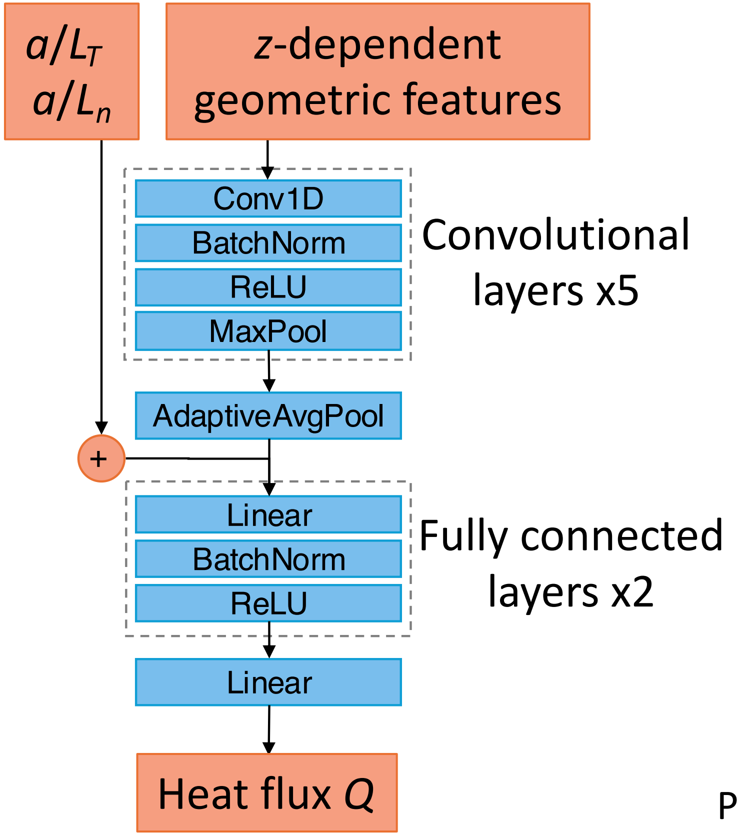
# Raw features should *not* be directly fed to classical regression or fully-connected neural network, since model should be translation-invariant

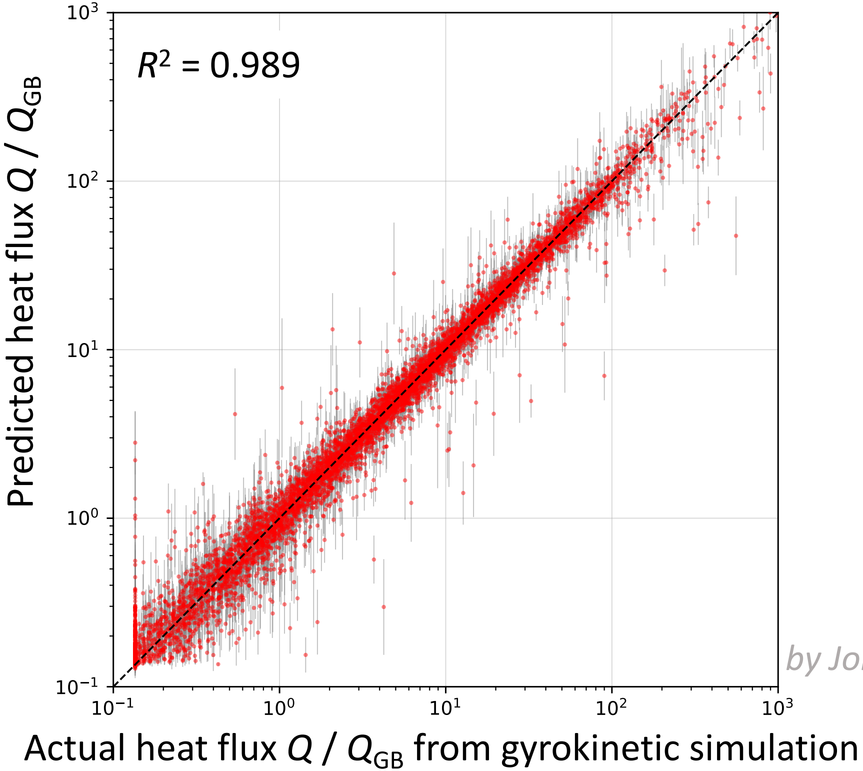
- GK equation, hence heat flux, is invariant under periodic translation of the raw features in $z$.
- Similar to computer vision, where convolutional neural networks give approximate translation-invariance.



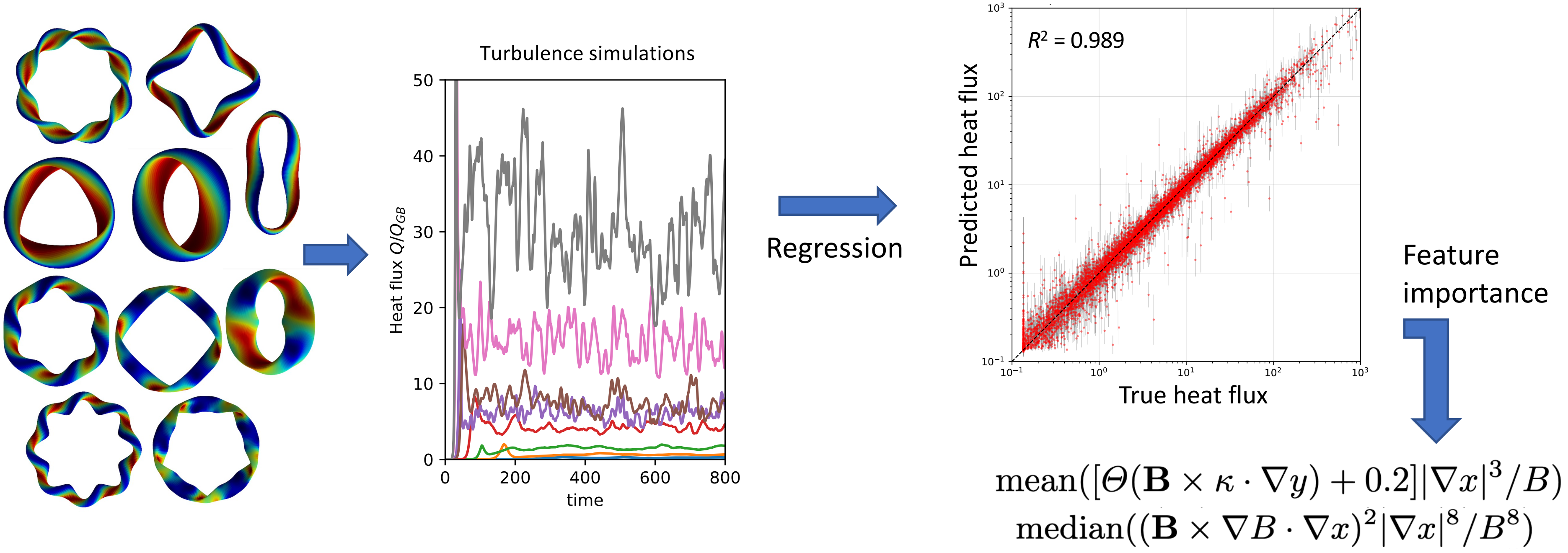
# Convolutional neural networks give accurate prediction of the turbulence




*by Jong Choi, ORNL*

Prediction in 0.001 sec for single network, 0.1 sec for ensemble



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

# 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 |
| --- | --- |
| $\text{variance}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\lvert\nabla x\rvert^2/B)$ | 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)

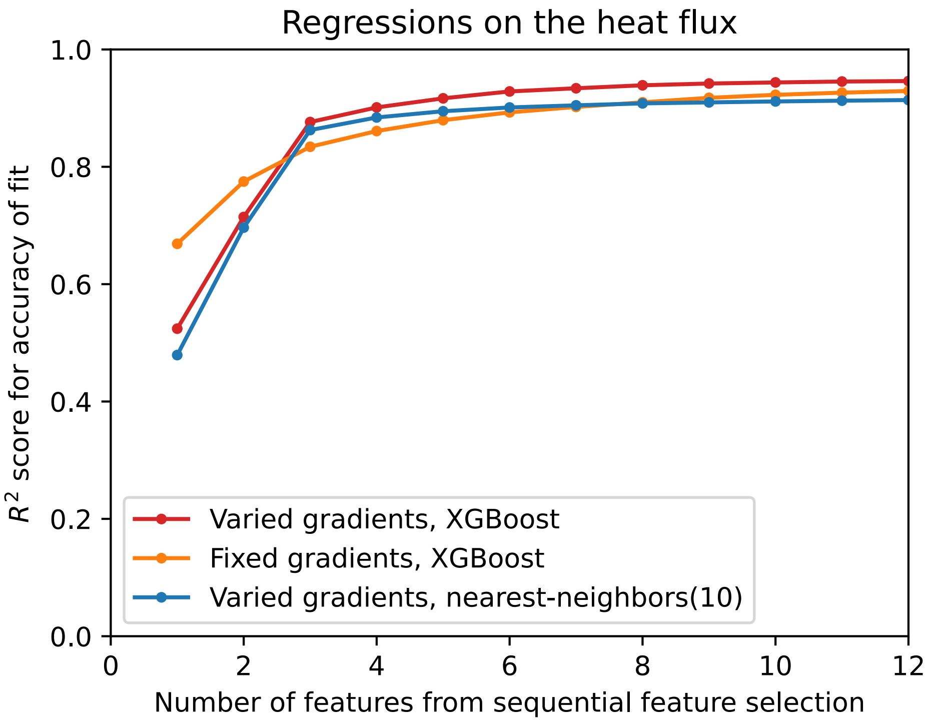
local temperature gradient in real space (to various powers)

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

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

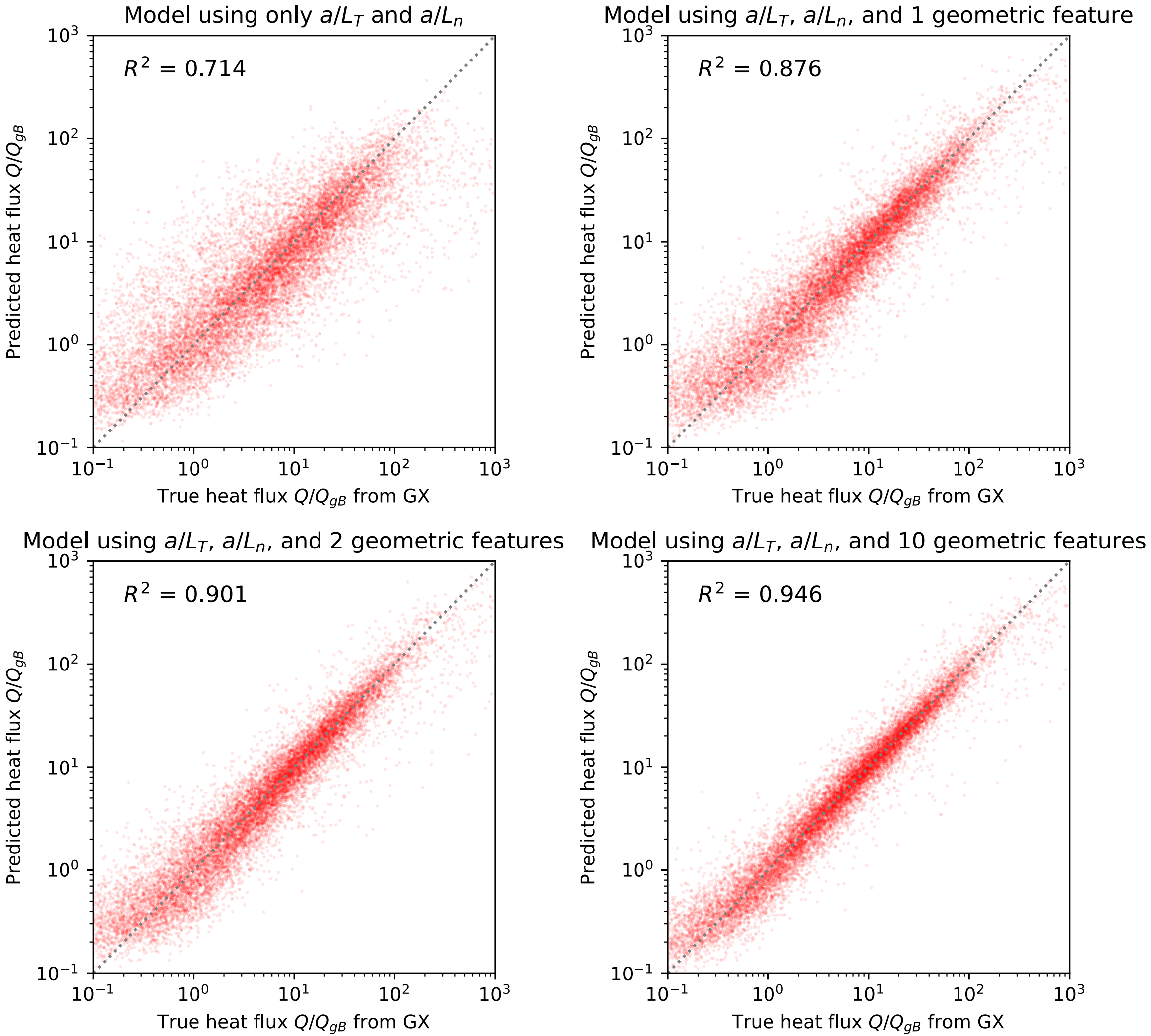
# Forward sequential feature selection: ~3 features can be almost as predictive as all features

## 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

# Most important features from sequential feature selection

**Regression on heat flux**

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

**Classification (stability vs instability)**

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

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

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

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

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

*Sequential feature selection step 3*

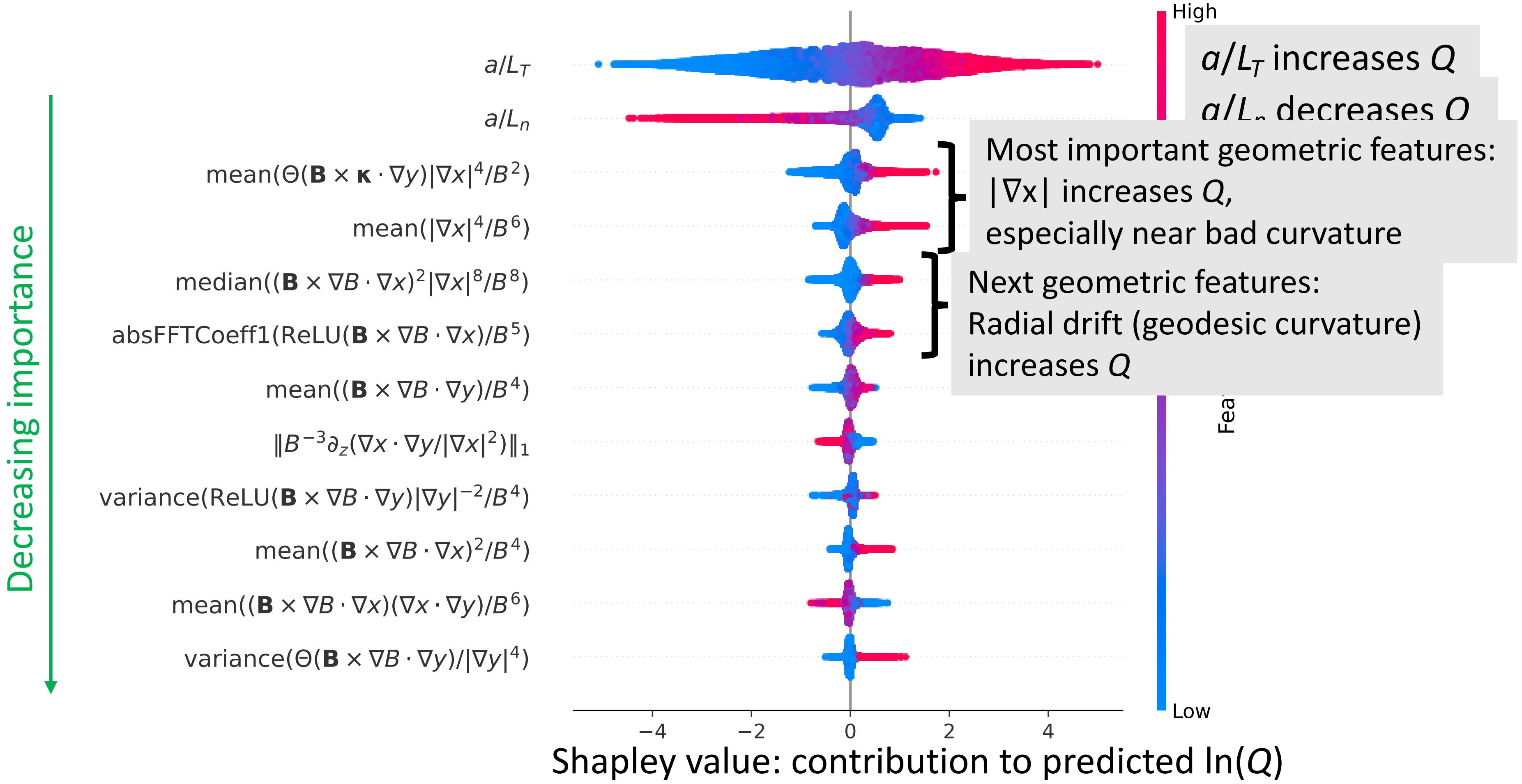
| Feature | $R^2$ |
|---|---|
| $\text{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\|\nabla x\|^4/B^2)$ | 0.876 |
| $\text{mean}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\|\nabla x\|^4/B)$ | 0.874 |
| $\text{variance}(\Theta(\mathbf{B} \times \kappa \cdot \nabla y)\|\nabla x\|^2/B)$ |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

*Sequential feature selection step 4*

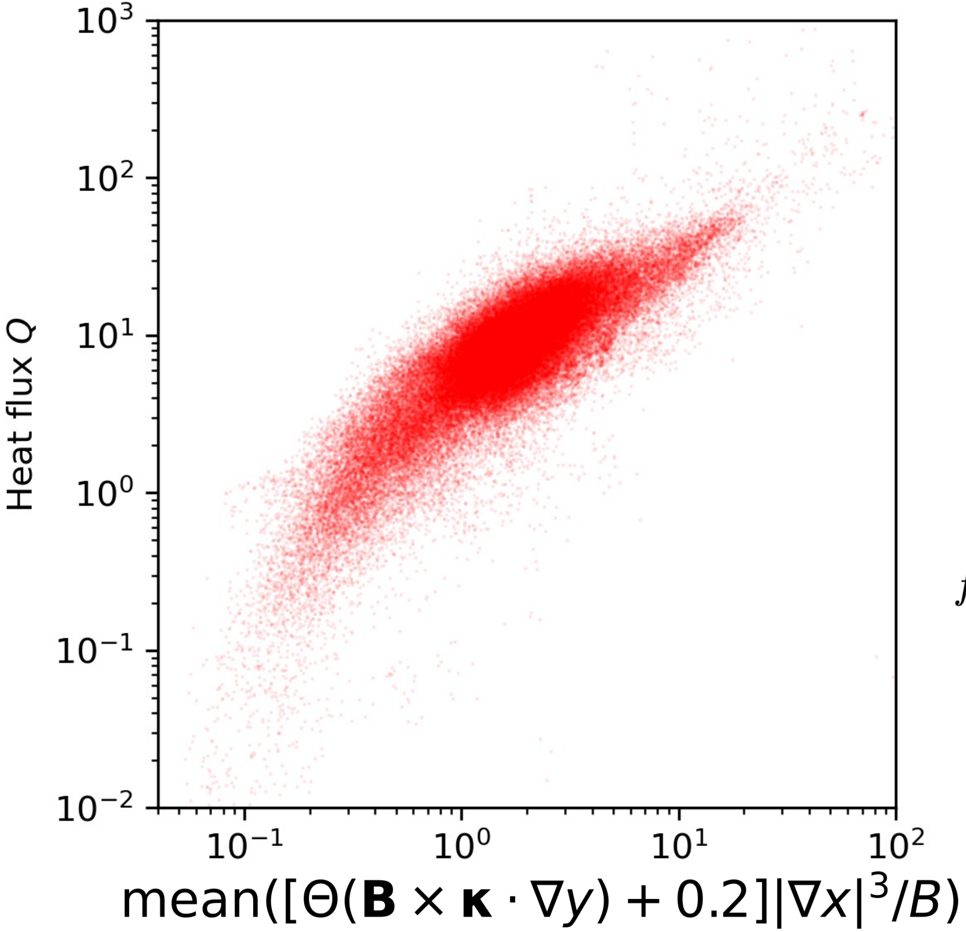
| Feature | $R^2$ |
|---|---|
| $\text{median}((\mathbf{B} \times \nabla B \cdot \nabla x)^2\|\nabla x\|^8/B^8)$ | 0.901 |
| $\text{quantile0.75}(\text{ReLU}(-\mathbf{B} \times \nabla B \cdot \nabla x)^2\|\nabla x\|^8/B^6)$ | 0.901 |
| $\text{median}((\mathbf{B} \times \nabla B \cdot \nabla x)^2\|\nabla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

*Regression for the random-gradient dataset*

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



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



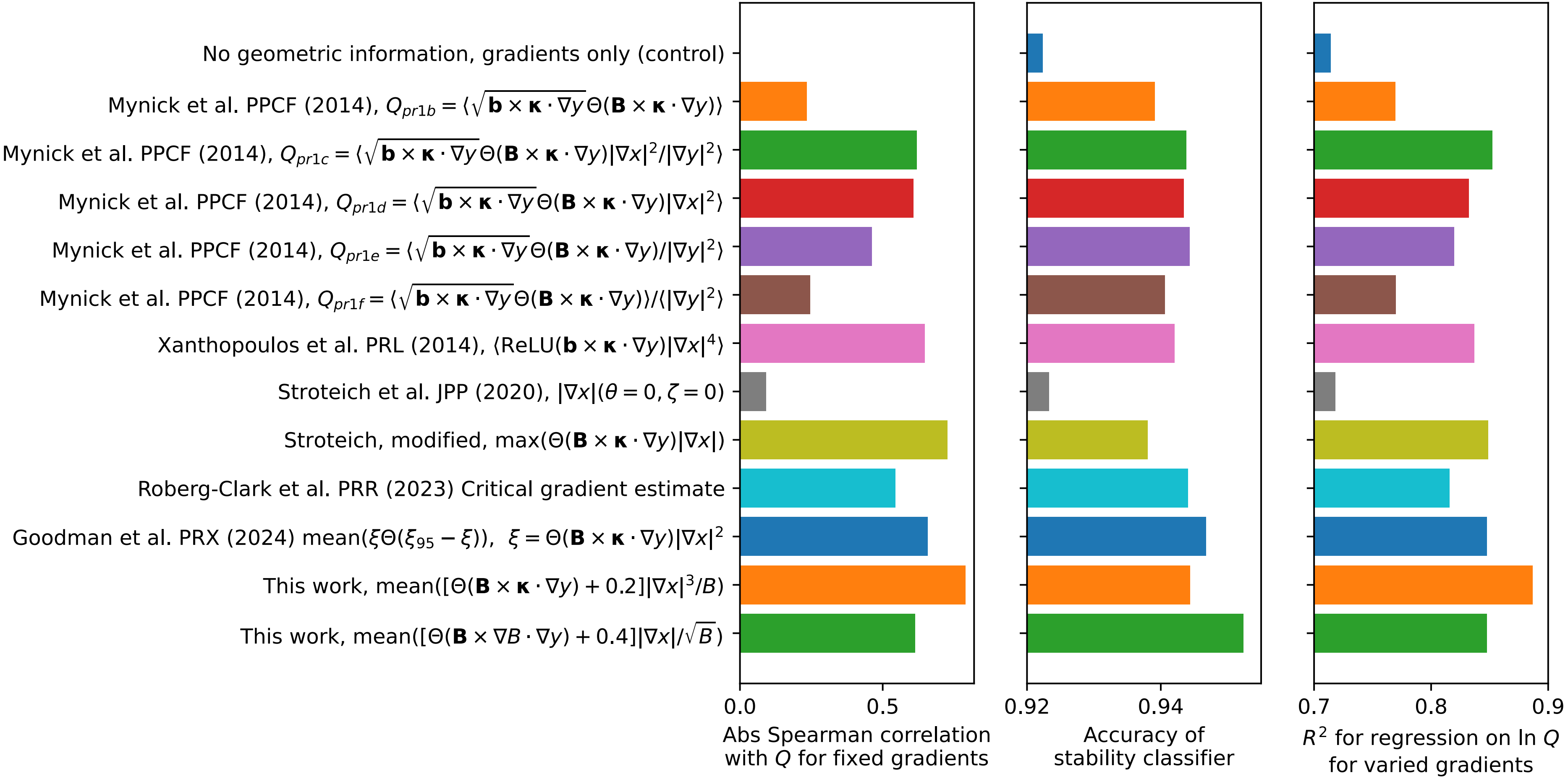

Fixed-gradient dataset.

No regression model used here.

Feature fine-tuned for stability classifier:

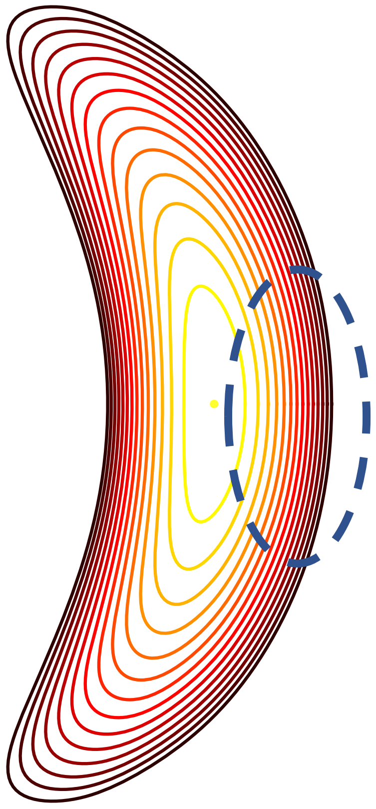
$$f_{\text{stab}} = \text{mean}([\Theta(\mathbf{B} \times \nabla B \cdot \nabla y) + 0.4] |\nabla x| / \sqrt{B})$$

# Previously proposed proxies can be tested



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

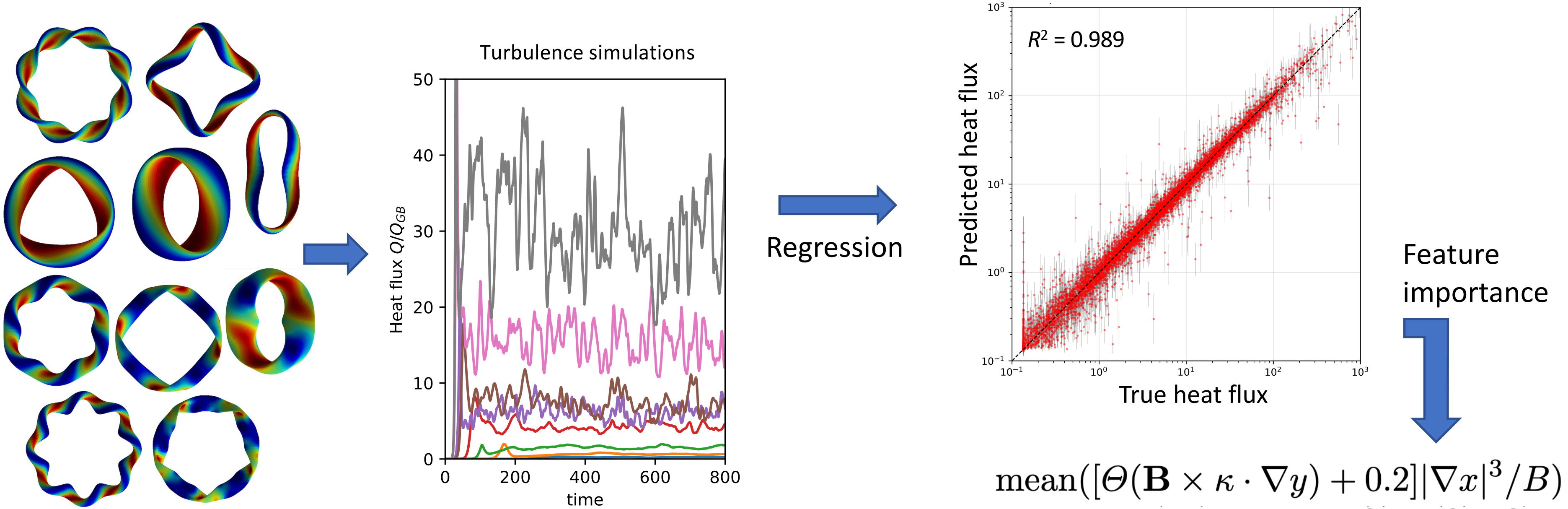
- Highest Spearman correlation at fixed gradients.
- Consistently the first geometric feature chosen in sequential feature selection:
  - In regression on the heat flux above the critical gradient
  - And in the classifier for stability vs instability (i.e. determines critical gradient)
  - Chosen by both XGBoost and nearest-neighbors.
- 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 will be released on Zenodo with the paper, so have a go at it!

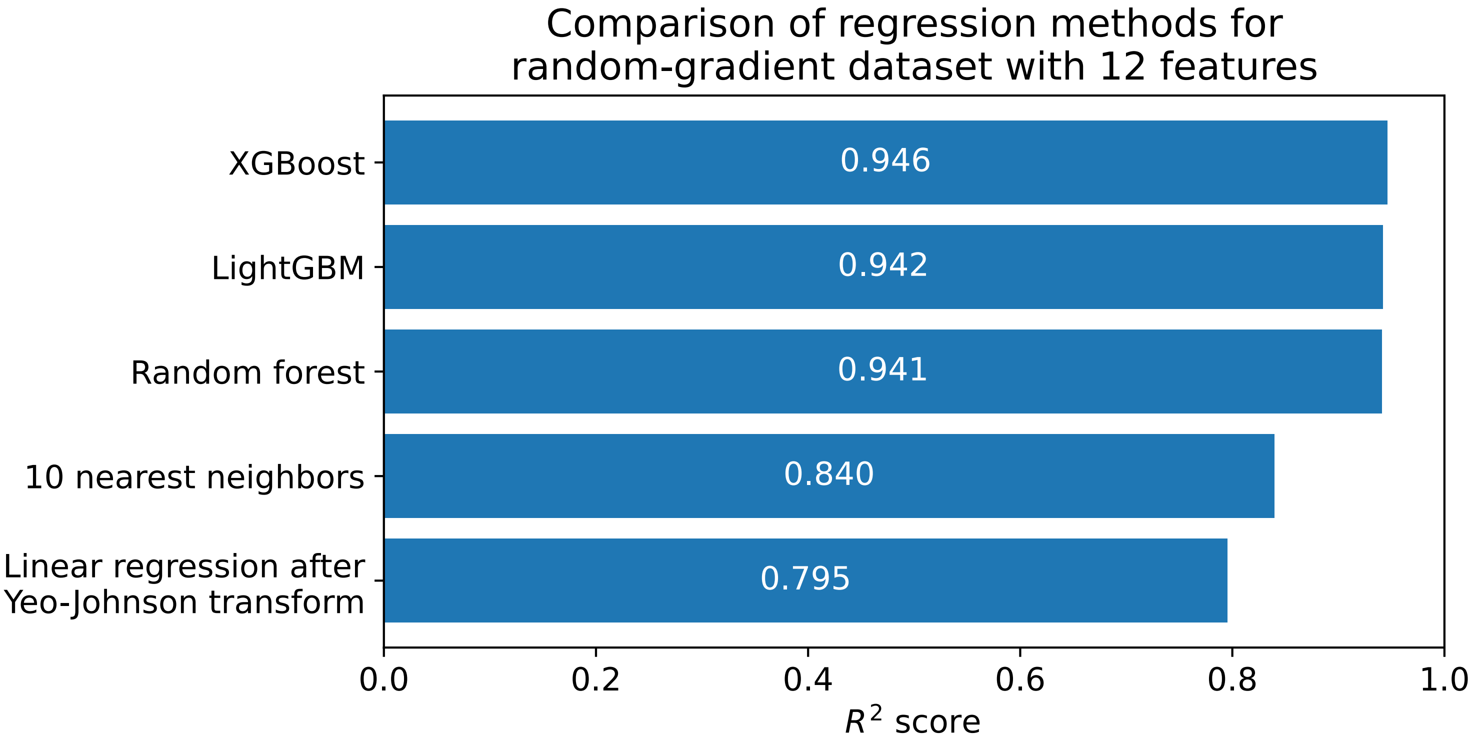



$$\mathrm{mean}([\Theta(\mathbf{B} \times \kappa \cdot \nabla y) + 0.2]|\nabla x|^3/B)$$

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

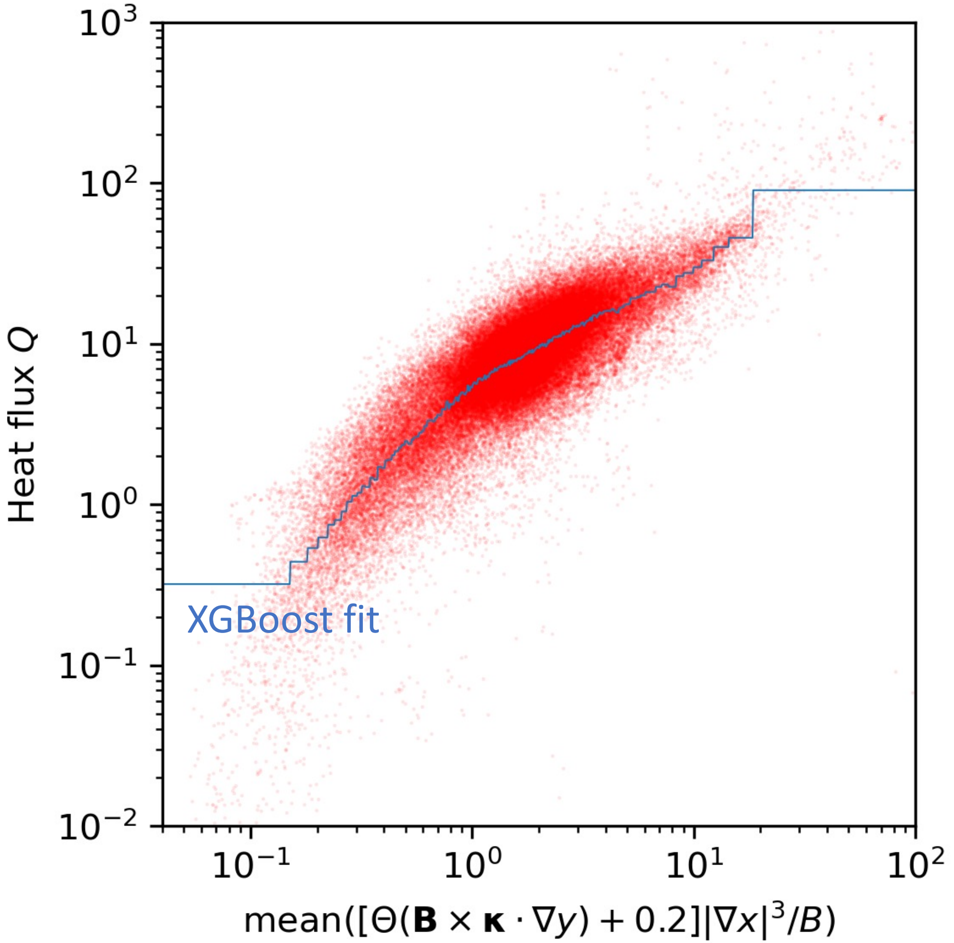
# Extra slides

# Other machine learning regression methods work also



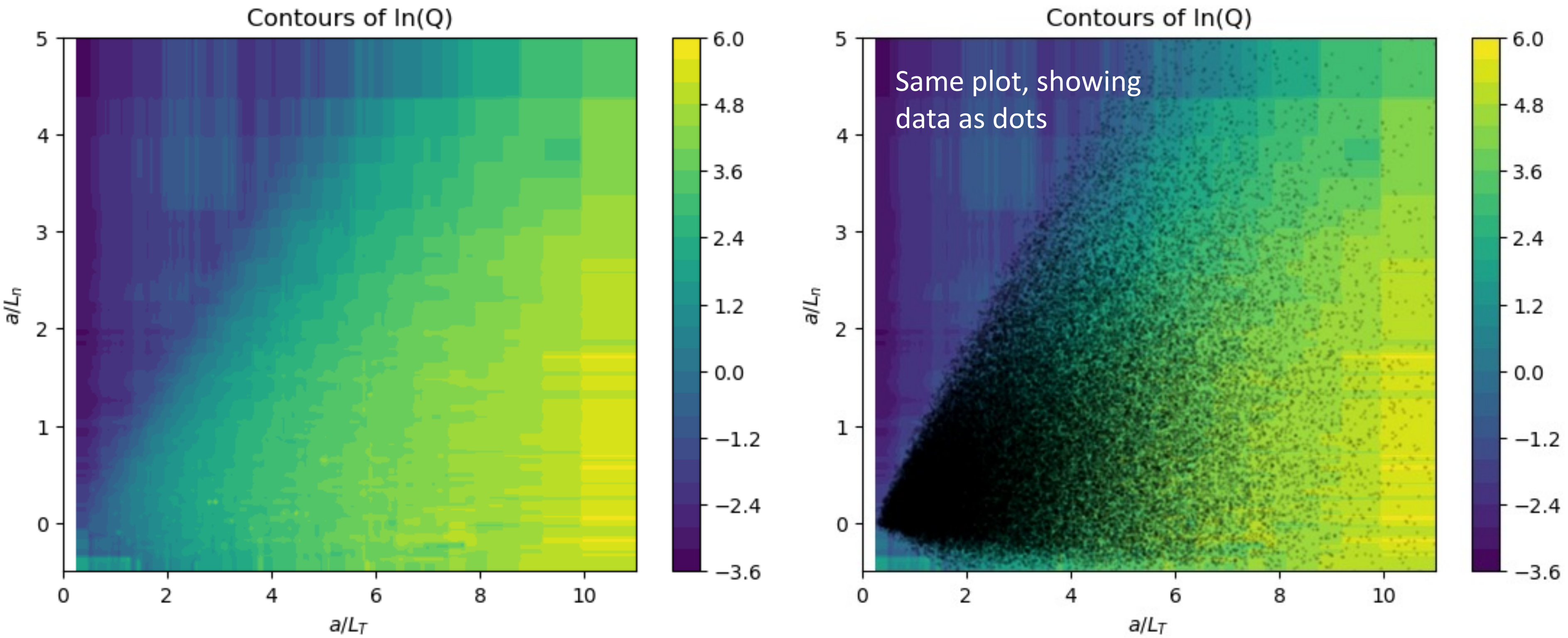

*All using $a/L_T$, $a/L_n$, and the top 10 geometric features selected via XGBoost*

# 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