Turbulence model reduction by deep learning

R.A. Heinonen and P.H. Diamond

CASS and Department of Physics
University of California, San Diego

Supported by the Department of Energy under Award Number DE-FG02-04ER54738

Outline

Introduction

Turbulent fluxes

- Turbulent flows characterized by strong fluctuations $\tilde{q}$ from mean $\langle q \rangle$ ($f = \mathbf{v}, T, n, \mathbf{B}, \mathbf{E}$, etc.)
- If fluctuations correlate, net transport in real space: e.g. convective flux $\Gamma_q = \langle \tilde{\mathbf{v}} \tilde{q} \rangle$
- Contain lots of interesting physics: control mean nonlinear dynamics, structure formation. Prediction/calculation critical to confinement problem
- How to model?
  1. Phenomenology (e.g. mixing length ansatz)
  2. Quasilinear theory
  3. Closure/renormalization
  4. Blindly simulate and make fancy color viewgraph(s)
- In this talk: explore alternative, *data-driven* method

Preview

- Impose only basic desired structure (e.g. locality) and symmetry constraints on flux model
- Use deep supervised learning to fill in the form of the dependencies from simulation
- As test of concept: apply to well-trodden ground (Hasegawa-Wakatani), and compare to analytics. Here, choose a particularly simple, interpretable ML model
- Recover existing theory, while finding some new features

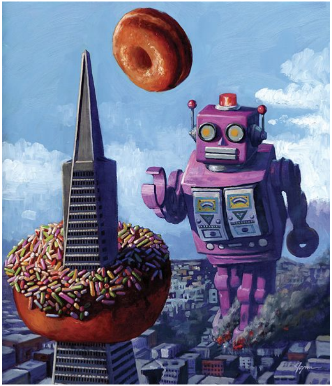


Figure Artist's conception of machine learning applied to the tokamak

Hasegawa-Wakatani

- Simplest realistic framework for understanding collisional drift wave turbulence

$$\frac{dn}{dt} = \alpha(\phi - n) + D\nabla^2 n$$
$$\frac{d\nabla^2\phi}{dt} = \alpha(\phi - n) + \mu\nabla^4\phi$$
$$\frac{d}{dt} \equiv \frac{\partial}{\partial t} + (\hat{z} \times \nabla\phi) \cdot \nabla$$

- $\alpha \equiv k_\parallel^2 T_e/(n_0 \eta \Omega_i e^2)$ "adiabaticity parameter," measures parallel electron response
- Want theory for *radial* transport (1D reduction)
- Averaging over symmetry directions ($\langle \cdots \rangle$) yields

$$\partial_t \langle n \rangle + \partial_x \Gamma = \text{dissipation}$$
$$\partial_t \langle \nabla^2 \phi \rangle - \partial_x^2 \Pi = \text{dissipation}$$

where $\Gamma = \langle \tilde{n}\tilde{v}_x \rangle$ (particle flux) and $\Pi = \langle \tilde{v}_x \tilde{v}_y \rangle$ (poloidal momentum flux or Reynolds stress)

Constructing a model

Locality

- Let's start by asking "what do we want our model to look like?"
- Standard simplifying assumption: locality in space and time. Assume local fluxes depend only on local means
- Might expect something like $\Gamma = -D\langle\varepsilon\rangle^{\alpha}\partial_x\langle n\rangle$ (turb. diffusion), $\Pi = -\chi\langle\varepsilon\rangle^{\beta}\partial_x^2\langle\phi\rangle + \dots$ (neg. viscosity)
- Sacrifices predictive power for interpretability

Feature selection I

- To use machine learning, need to decide what parameters flux should depend on
- Start with $\langle n \rangle, \partial_x \langle n \rangle, \partial_x^2 \langle n \rangle, \ldots,$ $\langle \phi \rangle, \partial_x \langle \phi \rangle, \partial_x^2 \langle \phi \rangle, \ldots, \langle \varepsilon \rangle, \partial_x \langle \varepsilon \rangle, \ldots$
- Need at least one measure ε of turbulence intensity. Theoretically convenient choice is turbulent PE $\langle (\tilde{n} - \nabla^2 \tilde{\phi})^2 \rangle$
- Must truncate chain of derivatives at some order. More derivatives $\rightarrow$ greater spatial nonlocality

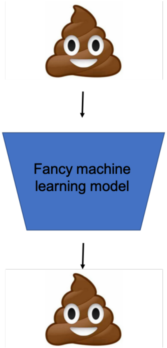


Figure Training a fancy state-of-the-art machine learning model on garbage inputs yields garbage predictions

Feature selection II

- Exact symmetries of HW are useful:
  1. Invariance under uniform shifts $n \to n + n_0$ and $\phi \to \phi + \phi_0$ eliminate dependence on $\langle n \rangle, \langle \phi \rangle$
  2. Invariance under Galilean boosts in y

$$\begin{cases} \phi & \to \phi + v_0 x \\ y & \to y - v_0 t \end{cases}$$

  eliminates dependence on ZF speed $\langle \partial_x \phi \rangle$ (not necessarily true in non-Markovian model!)

- Also exploit *approximate* poloidal symmetry so that $\langle \cdot \rangle$ is *zonal* average $\to$ reduction to 1D. Significant simplification, but restricts to ZF-dominated regimes

Feature selection III

- We choose a minimal set of parameters $N' = \partial_x \langle n \rangle, U = -\partial_x^2 \langle \phi \rangle, U', U'', \varepsilon$
- Anticipate that at least U'' needed to stabilize negative viscosity at small scales
- To close model, also will need dynamical equation for ε. Can derive from HW:

$$\partial_t \varepsilon + 2\varepsilon(\Gamma - \partial_x \Pi)(N' + U') = -\gamma\varepsilon - \gamma_{NL}\varepsilon^2$$

Linear damping accounts for stability threshold. Nonlinear damping (transfer to dissipation) compensates for neglected eddy-eddy term. Spreading term $\langle \tilde{v}_x (\tilde{n} - \nabla^2 \tilde{\phi})^2 \rangle$ neglected

Deep learning

Methods

- Deep supervised learning
- Locality $\rightarrow$ good scaling. Each point in space and time treated on equal footing!
- Exploit 3 reflection symmetries $x \rightarrow -x, y \rightarrow -y$ and $\phi \rightarrow -\phi, n \rightarrow -n, x \rightarrow -x$ and $\phi \rightarrow -\phi, n \rightarrow -n, y \rightarrow -y$ for data augmentation. Symmetries enforce, e.g. $\Gamma \rightarrow -\Gamma$ under $N' \rightarrow -N'$ in absence of flow
- Each simulation thus yields $4N_t N_x$ data points

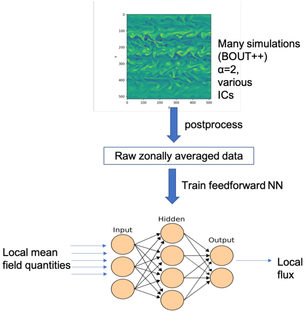


Figure Schematic of deep learning method

Deep neural networks 101

- Method of approximating arbitrary nonlinear functions. We use simplest form: "multi-layer perceptron."
- Inputs $\mathbf{x}$ repeatedly transformed in each layer: $x_j^{(n+1)} = \sigma(W_{ij}^{(n)} x_i^{(n)} + b_j^{(n)})$ where σ is a nonlinear function ("activation")
- Weights $\mathbf{W}^{(n)}$, biases $\mathbf{b}$ are "trained" using SGD
- Bottom line: simply a proven choice of multivariate, fully nonlinear, nonparametric regression

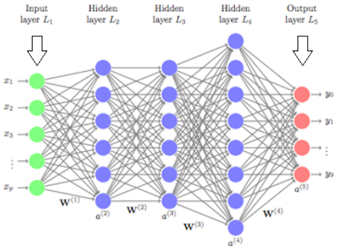


Figure Diagram of MLP, shamelessly stolen from the internet

Results

Particle flux

DNN learns a model roughly of the form (for small gradients)

$$\Gamma \simeq -D_n \varepsilon N' + D_U \varepsilon U'.$$

Large gradients: fluxes saturate. Diffusive term $\propto N'$ well-known, tends relax driving gradient. Second (non-diffusive) term is not so well-known, driven by vorticity gradient!

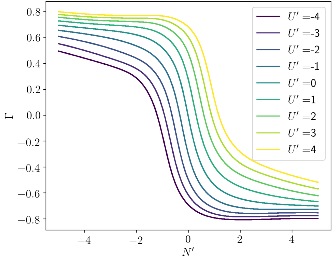


Figure Particle flux at constant ε as function of density and vorticity gradients

Derivation of nondiffusive term

Analytic treatment in $\alpha \to \infty$ limit reproduces nondiffusive term.
Need include frequency shift due to convection of mean vorticity.
In QLT:

$$\omega_{\mathbf{k}} = \frac{k_y}{1+k^2}(N' + U') + O(\alpha^{-2})$$

$$\gamma_{\mathbf{k}} = \frac{k_y^2}{\alpha(1+k^2)^3}(N' + U')(k^2 N' - U') + O(\alpha^{-2})$$

$$\begin{aligned}
\Gamma &= \mathrm{Re} \sum_{\mathbf{k}} -i k_y \tilde{n}_{\mathbf{k}} \tilde{\phi}^*_{\mathbf{k}} \\
&= \sum_{\mathbf{k}} \frac{-k_y^2 \partial_x n(\gamma_{\mathbf{k}} + \alpha) + \alpha k_y \omega_{r,\mathbf{k}}}{\omega_{r,\mathbf{k}}^2 + (\gamma_{\mathbf{k}} + \alpha)^2} |\tilde{\phi}_{\mathbf{k}}|^2 \\
&= \frac{1}{\alpha} \sum_{\mathbf{k}} -\frac{k_y^2}{1+k^2} \left(k^2 N' - U'\right) |\tilde{\phi}_{\mathbf{k}}|^2 + O(\alpha^{-2})
\end{aligned}$$

Comparison to theory (diffusive term)

Compare DNN result to theory result using spectrum centered at most unstable $\mathbf{k}$ for $U' = 0$

$$\varepsilon_{\mathbf{k}} = \frac{\langle \varepsilon \rangle}{2\pi^2 \Delta k_x \Delta k_y} \frac{1}{1 + k_x^2/\Delta k_x^2} \left(\frac{1}{1 + (k_y - \sqrt{2})^2/\Delta k_y^2} + \frac{1}{1 + (k_y + \sqrt{2})^2/\Delta k_y^2} \right)$$

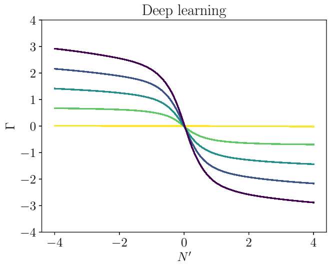


Figure Curves (at fixed $U = U' = U'' = 0$, and various ε) of Γ vs density gradient from DNN

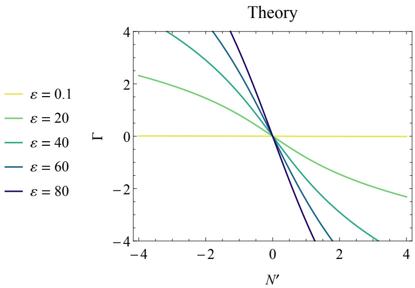


Figure Corresponding curves from QLT+ansatz with $\Delta k_x = \Delta k_y = 0.8$

Comparison to theory (nondiffusive term)

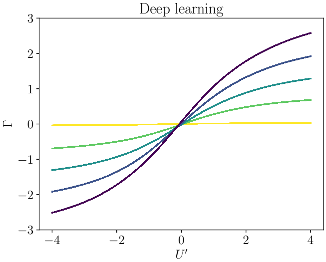


Figure Curves (at fixed $N' = U = U'' = 0$, and various ε) of Γ vs U' from DNN

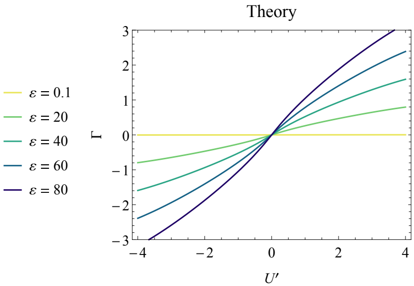


Figure Corresponding curves from QLT+ansatz with $\Delta k_x = \Delta k_y = 0.8$

Good agreement when N', U' are small!

Implications of nondiffusive term

- Neglected in literature, but coupling same order of magnitude (~ 0.5) that of usual N' term. DNN picks it out very clearly!
- Consequence: ZF can induce "staircase" pattern on profile. If $V_y = V_0 \sin(qx)$, U' term will contribute

$$\partial_t \langle n \rangle \sim -\frac{k_y^2 q^3 V_0 \langle \varepsilon \rangle}{\alpha (1 + k^2)^3} \cos(qx)$$

- Previous explanation for staircase is some form of bistability. This mechanism is distinct.

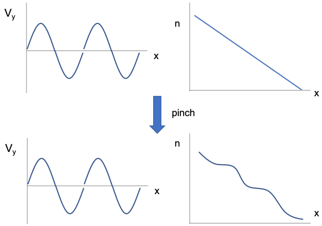


Figure Cartoon indicating how ZF may induce profile staircase via nondiffusive flux/pinch

What about the shear?

- Shear U is usually invoked as directly involved in suppression of turbulent transport
- We find that direct dependence on U is comparatively weak
- Suppression is $\lesssim 10\%$ for typical values of U
- Conclusion: for HW particle transport in 2D model, shear *gradient* more important than shear itself!

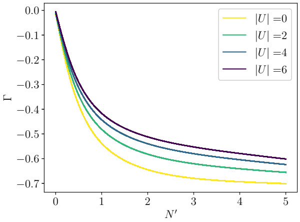


Figure U level curves of particle flux as function of N', at fixed $U'U''$, ε

Reynolds stress

- Learns model of Cahn-Hilliard form (leading order)

$$\Pi = \varepsilon(-\chi_1 U + \chi_3 U^3 - \chi_4 U'')$$

  with $\chi_1, \chi_3, \chi_4 > 0$
- $\partial_t U = \partial_x^2 \Pi \sim \chi_1 \varepsilon k^2 U$. Zonal flow generation by *negative viscosity* $\varepsilon\chi_1$
- Large U stabilized by nonlinearity $\propto U^3$, small scales by hyperviscosity χ_4
- Power law decay of Π with U at large U agrees with wave kinetic calculation

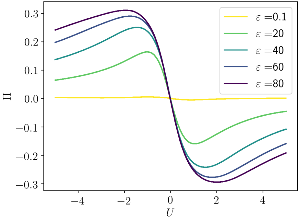


Figure: Reynolds stress as function of U, at fixed U', U''

Reynolds stress: hyperviscosity

Hyperviscous term, crucial for stability, has small coefficient. Sensitive test of method

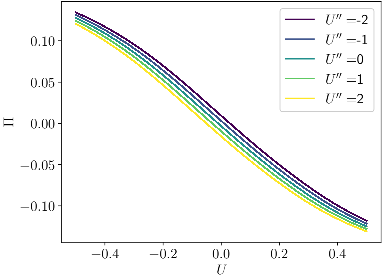


Figure U'' level curves of Reynolds stress as function of U, at fixed ε, U', N'

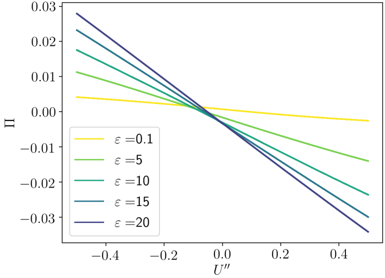


Figure ε level curves of Reynolds stress as function of U'', at fixed U, U', N''

Reynolds stress: gradient corrections

- How does Reynolds stress depend on N', U'? Not easy to calculate
- Learned dependence well-described by overall suppression factor $f \simeq 1/(1+0.04(N'+4U')^2)$, i.e. gradients generally reduce Reynolds stress
- Found to be crucial for stability of learned model. Kinks tend to form in flow in its absence

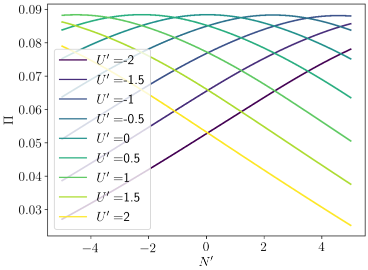
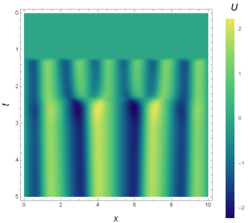


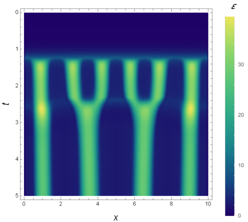
Figure Reynolds stress dependence on gradients at fixed ε, U, U''

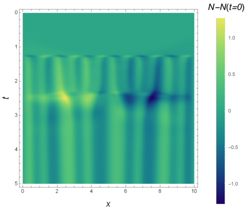
Discussion

Reduced 1-D model

Now have 3 coupled, **one-dimensional** mean field equations describing nonlinear turbulent dynamics. Construct expressions for Γ, Π capturing NN behavior, and numerically solve







Conclusions

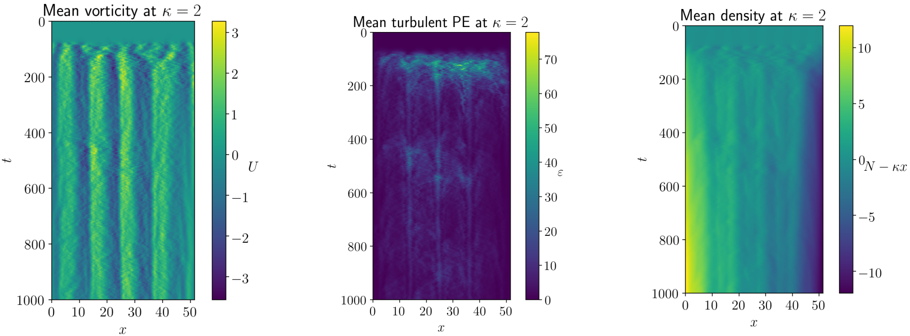
- Have verified, directly from simulation, analytic models for spontaneous ZF. CH is "best" local 1D model
- Identified significant vorticity-gradient-driven particle flux which may induce layering. Can distinguish from bistability e.g. by relative phase of U, N. Shearing effects weak
- Also find higher-order corrections which are harder to anticipate analytically (e.g. effect of shear on Γ, gradients on Π)
- Caveats:
  1. Need sufficient data. Serious issue for application to GK, experiment
  2. 1D reduction breaks down for strong turbulence due to vortex interactions, $\alpha \lesssim 1$ due to breakdown of ZF

Ideas for future

- No 1D reduction. Replace zonal average with 2D window average
- Relax locality assumption. Can include time derivatives as inputs, or extend to fully non-Markovian and/or spatially nonlocal models. Tradeoff is more predictive power at the expense of simplicity/interpretability
- This work essentially a second-order moment closure. Higher-order moments? Turbulence spreading? (interesting to note: applying this method to PE flux didn't work!)

Extra slides

Compare to zonally averaged 2D DNS



1D resembles simplified version of DNS. One key difference: 3-field model equivalent to taking stationary “best-fit” spectrum. Some system memory lost

Reynolds stress: intensity scaling

- Whereas learned Γ is essentially $\propto \varepsilon$, Π scaling with ε is nontrivial
- Learned exponent is 1 for small intensity, close to zero for large intensity
- Jibes with intuition from strong turbulence theory

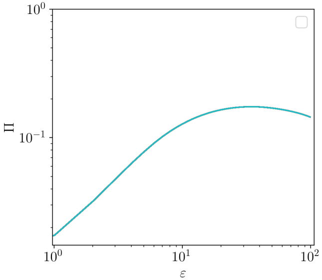


Figure Reynolds stress dependence on gradients at fixed ε, U, U''