Using transfer learning & backscattering analysis to build *stable, generalizable, data-driven* subgrid-scale (SGS) models: A 2D turbulence test case

Pedram Hassanzadeh  
Work with: Yifei Guan, Ashesh Chattopadhyay & Adam Subel  
*Rice University*

March 2021
Climate System:
spatio-temporal, *multi-scale*, *multi-physics*, high-dimensional & chaotic …

**X**: large/slow-scale variables
The main variables of interest

**Y**: small/fast-scale variables
Influence the spatio-temporal variability of *X*
Traditional approach:
Coarse-resolution numerical solver + physics-based subgrid-scale (SGS) model

Large-scale processes

\[ \dot{X} = F(X) \]

solved numerically at \( O(100) \) km resolutions

Closure for SGS processes

\[ Y = P(X) \]
ML-based approach:
Coarse-resolution numerical solver + data-driven subgrid-scale (SGS) model

Large-scale processes

\[ \dot{X} = F(X) \]

solved numerically at \( O(100) \)km resolutions

Data-driven closure for SGS processes

\[ Y = NN(X) \]

http://www-personal.umich.edu/~cjablono/
Using ML for weather/climate modeling: Questions, challenges & opportunities

• Best ways to use ML?
• How to choose the ML method?
• Dealing with poor (high-quality) data regimes
• Incorporating physics/PDEs’ properties
• Interpretability
• Generalization (i.e., extrapolation)
• Instability: blow-up in coupled (ML+numerical solver) models
Test case: 2D Turbulence

\[ \frac{\partial \omega}{\partial t} + J(\omega, \psi) = \frac{1}{\text{Re}} \nabla^2 \omega \]

\[ \nabla^2 \psi = -\omega \]
Large-Eddy Simulation (LES)

\[
\frac{\partial \omega}{\partial t} + J(\omega, \psi) = \frac{1}{\text{Re}} \nabla^2 \omega
\]

\[
\frac{\partial \bar{\omega}}{\partial t} + J(\bar{\omega}, \bar{\psi}) = \frac{1}{\text{Re}} \nabla^2 \bar{\omega} + \left[ J(\bar{\omega}, \bar{\psi}) - J(\omega, \psi) \right]
\]

\[\Pi: \text{SGS term}\]

Re=32000
DNS grid = 2048 x 2048, time step = \(\Delta t\)
LES grid = 256 x 256, time step = 10\(\Delta t\)

physics-based parameterization: Smagorinsky’s model (1963) \(\Pi = \nu_c \nabla^2 \bar{\omega}\)

data-driven parameterization (DD-P): \(\Pi = \text{NN}(\bar{\omega}, \bar{\psi})\)

Guan, Chattopadhyay, Subel & Hassanzadeh, Stable a posteriori LES of 2D turbulence with convolutional neural networks: backscattering analysis and generalization to higher Re via transfer learning, under review arXiv: 2102.11400
Major shortcoming of many physics-based models: Only diffusive, not accounting for backscattering

\[ T = \Pi \nabla^2 \bar{\omega} \]

\[ T^{DSMAG} = \nu_e \nabla^2 \bar{\omega} \nabla^2 \bar{\omega} \geq 0 \]
(Dynamic Smagorinsky, Germano et al. 1991)

\[ T^{FDNS} \quad T^{DMSAG} \]

Forward transfer (form resolved to unresolved scales)

Backscattering (form unresolved to resolved scales)
Non-local DD-P using CNNs

Training dataset:
From 7 DNS runs started from random initial conditions

Validation dataset:
From 3 DNS runs started from random initial conditions

Testing dataset:
From 5 DNS runs started from random initial conditions

10 layers (64 filters, 5 x 5) + ReLU + no pooling/upsampling
A priori (offline) test of DD-P

"online ≠ offline"
Stephan Rasp

<table>
<thead>
<tr>
<th></th>
<th>SMAG</th>
<th>DSMAG</th>
<th>ANN</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation coefficient $c$</td>
<td>0.55</td>
<td>0.55</td>
<td>0.86</td>
<td>0.93</td>
</tr>
</tbody>
</table>
Stability of a posteriori (coupled) LES model?

A priori accuracy of the CNN-based DD-P & the fate of coupled LES run as a function of the number of training samples, $N$

<table>
<thead>
<tr>
<th>$N$</th>
<th>500</th>
<th>1000</th>
<th>10000</th>
<th>30000</th>
<th>50000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_T &gt; 0$: diffusion</td>
<td>0.84</td>
<td>0.89</td>
<td>0.91</td>
<td>0.93</td>
<td>0.96</td>
</tr>
<tr>
<td>$c_T &lt; 0$: backscatter</td>
<td>0.56</td>
<td>0.71</td>
<td>0.82</td>
<td>0.85</td>
<td>0.90</td>
</tr>
<tr>
<td>Fate</td>
<td>Unstable</td>
<td>Unstable</td>
<td>Unstable</td>
<td>Stable</td>
<td>Stable</td>
</tr>
</tbody>
</table>

- Backscattering is harder to learn data drivenly when the training set is small
- Speculation: Disproportionally low accuracy for backscattering is the reason for instabilities
Accuracy of *a posteriori* (online) LES with DD-P

FDNS  
LES-CNN

LES-ANN  
LES-SMAG

Relative RMSE of $\overline{\omega}$

Time/$\tau$

- CNN
- DSMAG
- SMAG
- ANN
Accuracy of a posteriori (online) LES with DD-P

\[ \hat{E}(k) \text{ spectrum} \]

\[ \frac{\omega}{\sigma_\omega} \text{ PDF of vorticity} \]

- DNS
- FDNS
- CNN
- CNN w/o backscatter
- DSMAG
- SMAG
- ANN
DD-P does not generalize to higher $Re$

LES resolution: 256 x 256
Generalization to higher *Re* via transfer learning


Layers 1–8: Fixed (trained with *N* samples from *Re*)  
Re-train with 0.01*N* samples from higher *Re*
Generalization to higher $Re$ via transfer learning

Test on $Re = 8000$

Test on $Re = 32000$

Test on $Re = 64000$

LES resolution: 256 x 256
Generalization to higher $Re$ & different grid resolution via transfer learning + auto-encoder

<table>
<thead>
<tr>
<th></th>
<th>$Re$</th>
<th>$N$</th>
<th>grid resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base CNN</td>
<td>8000</td>
<td>50,000</td>
<td>256 x 256</td>
</tr>
<tr>
<td>Transfer learned CNN</td>
<td>128000</td>
<td>500</td>
<td>512 x 512</td>
</tr>
</tbody>
</table>

Test on $Re = 128000$
1D Stochastically forced Burgers turbulence

\[
\frac{\partial u}{\partial t} + \frac{1}{2} \frac{\partial uu}{\partial x} = \frac{1}{\text{Re}} \frac{\partial^2 u}{\partial x^2} + f(x, t)
\]

\[
\frac{\partial \bar{u}}{\partial t} + \frac{1}{2} \frac{\partial \bar{u}\bar{u}}{\partial y} = \frac{1}{\text{Re}} \frac{\partial^2 \bar{u}}{\partial y^2} + \bar{f} + \Pi(y)
\]

\[\Pi(y) = \text{ANN}(\bar{u})\]

- Stable \textit{a posteriori} LES after data augmentation
- Generalization to 10x Re using transfer learning

Subel, Chattopadhyay, Guan & Hassanzadeh, \textit{Data-driven subgrid-scale modeling of forced Burgers turbulence using deep learning with generalization to higher Reynolds numbers via transfer learning}, Physics of Fluids (2021)
Stable, accurate & generalizable SGS modeling for LES

Takeaway:
- Stability: *might* require large training sets
- Transfer learning: large training sets required only from a base system

- Reduce the required size of the training set
  - Data augmentation
  - Include physics constraints

- Better understanding of the relationship between a priori accuracy & a posteriori stability
- Online training?
- Add memory to the SGS model
- Further explore the power of transfer learning (e.g., between setups)
- More complex test cases
Papers on data-driven forecasting
http://pedram.rice.edu/publications/


Chattopadhyay A., Mustafa M., Hassanzadeh P., Bach E. & Kashinath K., Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving deep spatial transformers, under review at *Geoscientific Model Development*
