Small closure models for large multiscale problems

Wouter Edeling, Daan Crommelin
Scientific Computing Group

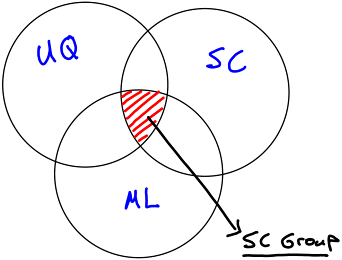
CWI Scientific Meeting, June 17th, 2022.

About me

- Aerospace Engineering TU Delft: MSc.
- Arts et Métiers ParisTech & TU Delft: joint-PhD.
- Stanford University: postdoc.
- CWI: postdoc & tenure track.

About me

- Aerospace Engineering TU Delft: MSc.
- Arts et Métiers ParisTech & TU Delft: joint-PhD.
- Stanford University: postdoc.
- CWI: postdoc & tenure track.



Topic: turbulent flow

What is turbulent flow?

Topic: turbulent flow

What is turbulent flow?

- Not laminar.

Topic: turbulent flow

What is turbulent flow?

- Not laminar.
- Unsteady.

Topic: turbulent flow

What is turbulent flow?

- Not laminar.
- Unsteady.
- Mixing.

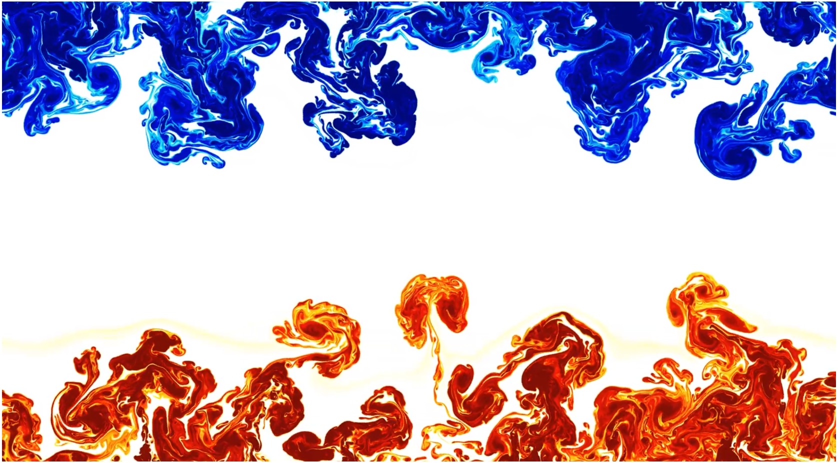
Topic: turbulent flow

What is turbulent flow?

- Not laminar.
- Unsteady.
- Mixing.
- <u>Multiscale</u>.

Numerical simulation

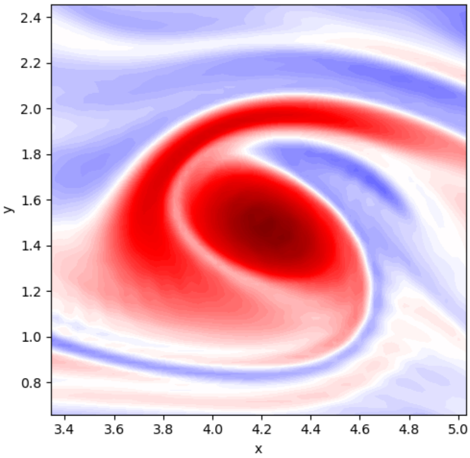
Numerical simulation resolving all spatial & temporal scales:



Credit: turbulence team:
https://www.youtube.com/watch?v=OM0l2YPVMf8

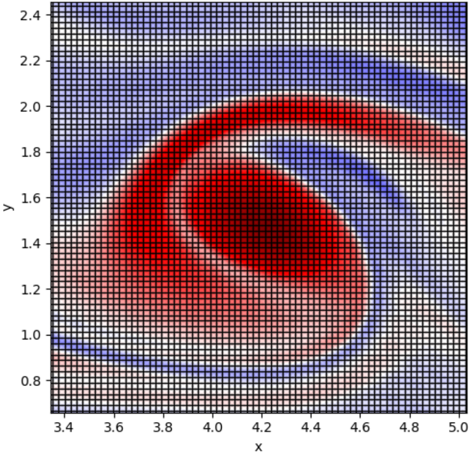
Discretization

- Numerical simulation = discretization: $\omega(x, y) \to \omega^h(x, y)$



Discretization

- Numerical simulation = discretization: $\omega(x, y) \to \omega^h(x, y)$



Solve equations on each point of a fine mesh.

Filtering

- Problem: multi-scale nature:
  $\rightarrow$ required mesh resolution (often) much too large.

Filtering

- Problem: multi-scale nature:
  $\rightarrow$ required mesh resolution (often) much too large.

- Engineering solution:
  $\rightarrow$ decompose solution $\boxed{\omega = \overline{\omega} + \omega'}$.
  $\rightarrow$ only solve for large scales $\overline{\omega}$.

Filtering

- Problem: multi-scale nature:
  $\rightarrow$ required mesh resolution (often) much too large.

- Engineering solution:
  $\rightarrow$ decompose solution $\boxed{\omega = \overline{\omega} + \omega'}$.
  $\rightarrow$ only solve for large scales $\overline{\omega}$.

- How to get $\overline{\omega}$?
  $\rightarrow$ Use filter $\boxed{\overline{\omega} = P\omega}$

Filtering

- Governing equations:

$$\frac{\partial \omega}{\partial t} + J(\psi, \omega) = \nu \nabla^2 \omega + \mu \left(F - \omega \right)$$
$$\nabla^2 \psi = \omega.$$

Filtering

- Governing equations:

$$\frac{\partial \omega}{\partial t} + J(\psi, \omega) = \nu \nabla^2 \omega + \mu (F - \omega)$$
$$\nabla^2 \psi = \omega.$$

- Apply filter:

$$\frac{\partial \bar{\omega}}{\partial t} + \overline{J\left(\bar{\psi}, \bar{\omega}\right)} = \nu \nabla^2 \bar{\omega} + \mu (F - \bar{\omega}) - \overline{r},$$
$$\nabla^2 \bar{\psi} = \bar{\omega}.$$

Filtering

- Governing equations:

$$\frac{\partial \omega}{\partial t} + J(\psi, \omega) = \nu \nabla^2 \omega + \mu (F - \omega)$$
$$\nabla^2 \psi = \omega.$$

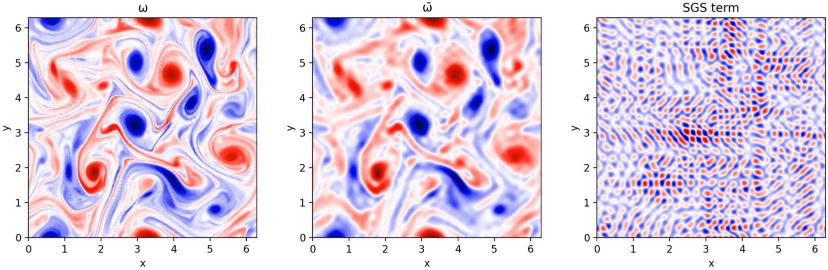
- Apply filter:

$$\frac{\partial \bar{\omega}}{\partial t} + \overline{J(\bar{\psi}, \bar{\omega})} = \nu \nabla^2 \bar{\omega} + \mu (F - \bar{\omega}) - \bar{r},$$
$$\nabla^2 \bar{\psi} = \bar{\omega}.$$

- Sub-Grid Scale (SGS) term $\bar{r}$ appears.

Filtering

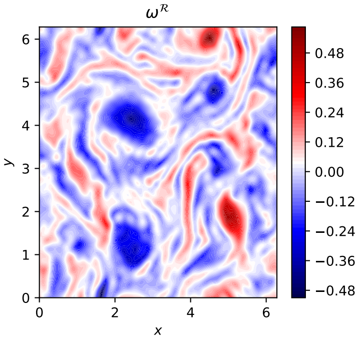
- Solving both equations side by side:

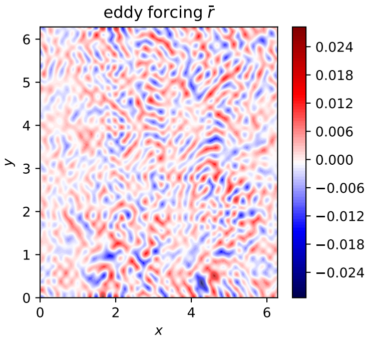


Closing governing equations

- Problem: SGS model is unknown / unclosed $\overline{r} = \overline{r}(\omega, \psi)$.
  $\rightarrow \overline{r}$ must be modelled

$$\frac{\partial \bar{\omega}}{\partial t} + \overline{J\left(\bar{\psi}, \bar{\omega}\right)} = \nu \nabla^2 \bar{\omega} + \mu \left(F - \bar{\omega}\right) - \overline{r},$$
$$\nabla^2 \bar{\psi} = \bar{\omega}.$$

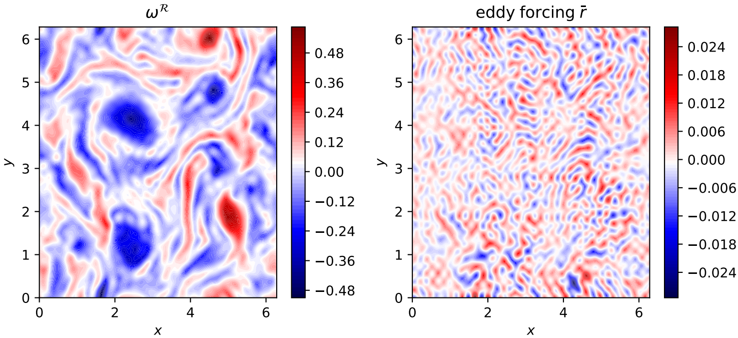




Closing governing equations

- Problem: SGS model is unknown / unclosed $\bar{r} = \bar{r}(\omega, \psi)$.
  $\rightarrow \bar{r}$ must be modelled

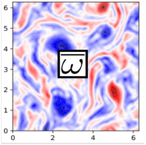
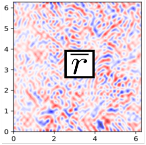
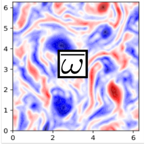
$$\frac{\partial \bar{\omega}}{\partial t} + \overline{J\left(\bar{\psi}, \bar{\omega}\right)} = \nu \nabla^2 \bar{\omega} + \mu \left(F - \bar{\omega}\right) - \bar{r},$$
$$\nabla^2 \bar{\psi} = \bar{\omega}.$$



- Goal: Learn $\bar{r}$ from 256×256 simulation.

Question

- What should we learn from data?

	state size	SGS training data size	QoI size
traditional training	64×64  →	64×64  →	$Q(\overline{\omega}) \in \mathbb{R}^{64 \times 64}$
reduced training	64×64  →	 →	$q(\overline{\omega}) \in \mathbb{R}^{d}$ $d \ll 64 \times 64$

Assumptions

1 There are d global QoI:

$$q_i(t) = \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} \int_0^{2\pi} f_i(\bar{\omega}, \bar{\psi}; x, y, t)\, \mathrm{d}x\mathrm{d}y, \quad i = 1, \cdots, d.$$

Assumptions

1 There are d global QoI:

$$q_i(t) = \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} \int_0^{2\pi} f_i(\bar{\omega}, \bar{\psi}; x, y, t)\, \mathrm{d}x\mathrm{d}y, \quad i = 1, \cdots, d.$$

2 Replace $\overline{r} = \overline{r(\psi, \omega)}$ with <u>reduced</u> SGS term $\underline{r}$:

$$\boxed{\underline{r} := \sum_{i=1}^{d} \tau_i(t) P_i(x, y, t)},$$

Assumptions

$$\boxed{\underline{r} := \sum_{i=1}^{d} \tau_i(t) P_i(x, y, t)},$$

Justified if:

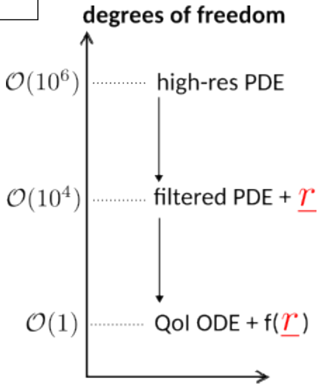
- $\underline{r}$ is 'just as good' as $\overline{r}$ for q_i.

Assumptions

$$\boxed{\underline{r} := \sum_{i=1}^{d} \tau_i(t) P_i(x, y, t)},$$

Justified if:

- $\underline{r}$ is 'just as good' as $\overline{r}$ for q_i.
- Must tie τ_i & P_i to q_i physics.



Compute effect of assumptions

- Derive q_i ODEs:

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = ... + \left(\frac{\partial f_i}{\partial \overline{\omega}}, \underline{r}\right) = ... + \sum_{j=1}^{d} \tau_j \left(\frac{\partial f_i}{\partial \overline{\omega}}, P_j\right)$$

- $(A, B) := \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} \int_0^{2\pi} AB \, \mathrm{d}x\mathrm{d}y.$

[1] Edeling, W., & Crommelin, D. (2020). Reducing data-driven dynamical subgrid scale models by physical constraints. Computers & Fluids, 201, 104470.

Compute effect of assumptions

- Derive q_i ODEs:

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \ldots + \left(\frac{\partial f_i}{\partial \bar{\omega}}, \underline{r}\right) = \ldots + \sum_{j=1}^{d} \tau_j \left(\frac{\partial f_i}{\partial \bar{\omega}}, P_j\right)$$

- $(A, B) := \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi}\int_0^{2\pi} AB \,\mathrm{d}x\mathrm{d}y.$
- Every q_i ODE has d SGS terms:

[1] Edeling, W., & Crommelin, D. (2020). Reducing data-driven dynamical subgrid scale models by physical constraints. Computers & Fluids, 201, 104470.

Compute effect of assumptions

- Derive q_i ODEs:

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = ... + \left(\frac{\partial f_i}{\partial \bar{\omega}}, \underline{r}\right) = ... + \sum_{j=1}^{d} \tau_j \left(\frac{\partial f_i}{\partial \bar{\omega}}, P_j\right)$$

- $(A, B) := \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} \int_0^{2\pi} AB \, \mathrm{d}x\mathrm{d}y.$
- Every q_i ODE has d SGS terms: remove $\sum_{j=1}^{d}$

[1] Edeling, W., & Crommelin, D. (2020). Reducing data-driven dynamical subgrid scale models by physical constraints. Computers & Fluids, 201, 104470.

Compute effect of assumptions

- Derive q_i ODEs:

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = ... + \left(\frac{\partial f_i}{\partial \bar{\omega}}, \underline{r}\right) = ... + \sum_{j=1}^{d} \tau_j \left(\frac{\partial f_i}{\partial \bar{\omega}}, P_j\right)$$

- $(A, B) := \left(\frac{1}{2\pi}\right)^2 \int_0^{2\pi} \int_0^{2\pi} AB \,\mathrm{d}x\mathrm{d}y.$
- Every q_i ODE has d SGS terms: remove $\sum_{j=1}^{d}$
- Orthogonality condition $\forall t$:

$$\boxed{\left(\frac{\partial f_i}{\partial \bar{\omega}}, P_j\right) = 0 \ \text{ if } \ i \neq j}$$

- Separate expansion for P_j and small linear solve [1].

[1] Edeling, W., & Crommelin, D. (2020). Reducing data-driven dynamical subgrid scale models by physical constraints. Computers & Fluids, 201, 104470.

Extract τ_i from data

- Due to orthogonality, q_i transport equation becomes:

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = ... + \left(\frac{\partial f_i}{\partial \bar{\omega}}, \underline{r}\right) = ... + \tau_i \left(\frac{\partial f_i}{\partial \bar{\omega}}, P_i\right)$$

- Goal: error $q_i^{ref}(t) - q_i(t) =: \Delta q_i$ is small $\forall t$ in training.

Extract τ_i from data

- Due to orthogonality, q_i transport equation becomes:

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = ... + \left(\frac{\partial f_i}{\partial \bar{\omega}}, \underline{r}\right) = ... + \tau_i \left(\frac{\partial f_i}{\partial \bar{\omega}}, P_i\right)$$

- Goal: error $q_i^{ref}(t) - q_i(t) =: \Delta q_i$ is small $\forall t$ in training.
- Assumption: τ_i depends upon Δq_i.

Extract τ_i from data

- Due to orthogonality, q_i transport equation becomes:

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = ... + \left(\frac{\partial f_i}{\partial \bar{\omega}}, \underline{r}\right) = ... + \tau_i \left(\frac{\partial f_i}{\partial \bar{\omega}}, P_i\right)$$

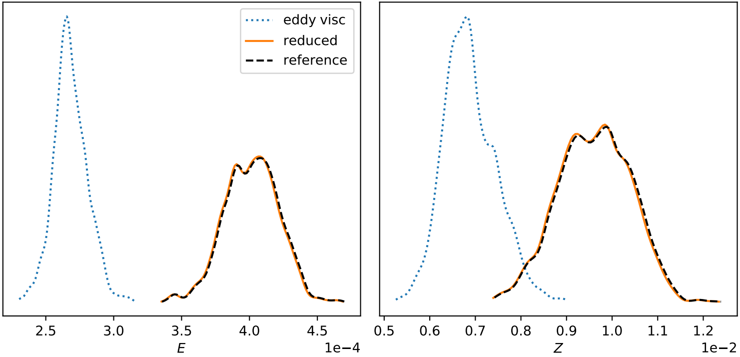
- Goal: error $q_i^{ref}(t) - q_i(t) =: \Delta q_i$ is small $\forall t$ in training.
- Assumption: τ_i depends upon Δq_i.
- Simply equate source term to Δq_i:

$$\tau_i \left(\frac{\partial f_i}{\partial \omega}, P_i\right) = \Delta q_i / T_i, \quad T_i = 1, \quad i = 1, \cdots, d$$

- Assumes $\tau_i \sim \Delta q_i$ + imposes linear relaxation towards reference.
- Δq_i is the only data we need.

Example results

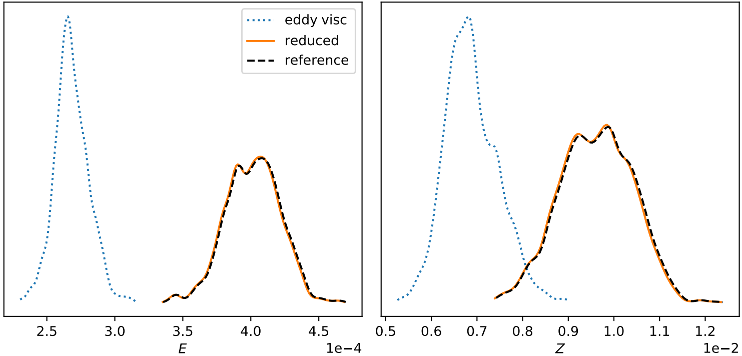
- $q_1 =$ energy E, $q_2 =$ enstrophy Z.



- $\underline{r}$ is 'just as good' as $\overline{r}$ for q_i.
  - $\rightarrow$ Number of unknowns reduced from 64^2 to 2.
  - $\rightarrow$ Training data size reduced by factor $64^2/2$.

Example results

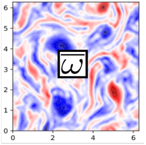
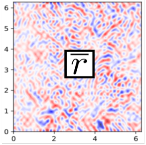
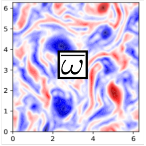
- q_1 = energy E, q_2 = enstrophy Z.



$$\underline{r} = -\frac{1}{2}\left[\frac{\Delta E}{S - E^2/Z}\right]\left(\Psi + \frac{E}{Z}\omega\right) + \frac{1}{2}\left[\frac{\Delta Z}{Z - E^2/S}\right]\left(\omega + \frac{E}{S}\Psi\right)$$

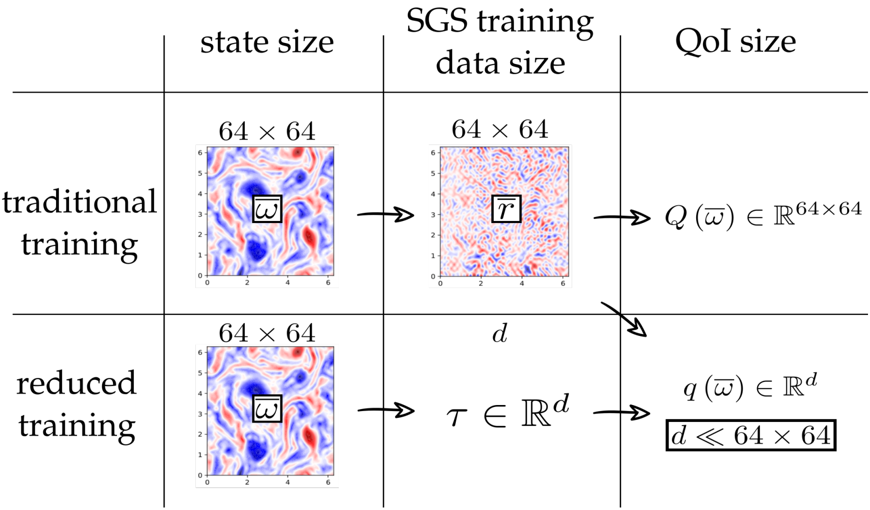
Question

- What should we learn from data?

	state size	SGS training data size	QoI size
traditional training	64×64  $\overline{\omega}$	64×64  $\overline{r}$	$Q(\overline{\omega}) \in \mathbb{R}^{64\times 64}$
reduced training	64×64  $\overline{\omega}$		$q(\overline{\omega}) \in \mathbb{R}^{d}$ $d \ll 64 \times 64$

Conclusion

- τ_i (or Δq_i)



Questions?

Edeling, W., & Crommelin, D. (2020). Reducing data-driven dynamical subgrid scale models by physical constraints. Computers & Fluids, 201, 104470.

Offline training

- Now: train ML model on reduced training data.
- Offline training: train e.g. ANN on $\Delta E, \Delta Z$ database.

Offline training

- Now: train ML model on reduced training data.
- Offline training: train e.g. ANN on $\Delta E, \Delta Z$ database.

Offline training

- Now: train ML model on reduced training data.
- Offline training: train e.g. ANN on $\Delta E, \Delta Z$ database.

Now we have a coupled PDE - ML system:

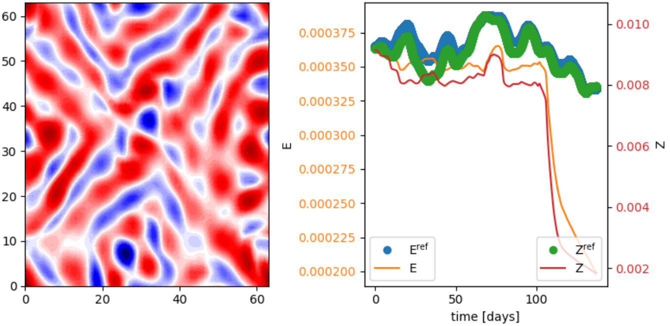
$$\frac{\partial \bar{\omega}}{\partial t} + \overline{J\left(\bar{\psi}, \bar{\omega}\right)} = \nu \nabla^2 \bar{\omega} + \mu \left(F - \bar{\omega}\right) - \tau_1(\widetilde{\Delta E}) P_1 - \tau_2(\widetilde{\Delta Z}) P_2,$$

$$\nabla^2 \bar{\psi} = \bar{\omega}.$$

$$[\widetilde{\Delta E}, \widetilde{\Delta Z}] = ANN(X_1, \cdots, X_7)$$

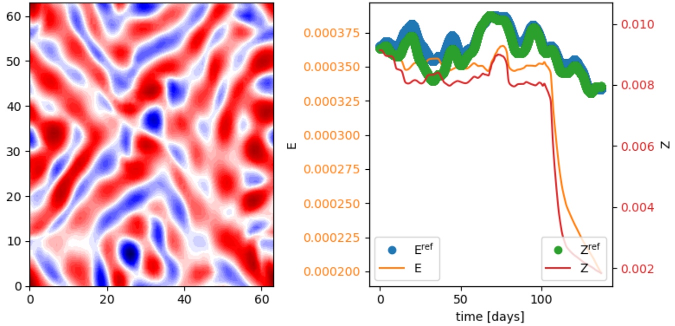
Prediction with offline surrogate

- Can become unstable:



Prediction with offline surrogate

- Can become unstable:



- Why?: ANN was not trained not to operate in a two-way coupled modelling environment.
- Other authors reported similar issues.

Online training

- online training while ANN is coupled to PDE.

[2] Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1. 0). Geoscientific Model Development, 13(5), 2185-2196.

[3] Sahoo, D. et al, (2017). Online deep learning: Learning deep neural networks on the fly. arXiv preprint arXiv:1711.03705.

Online training

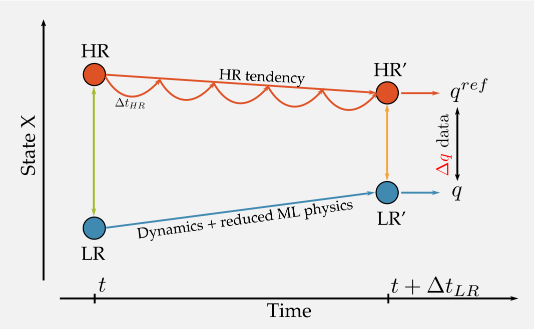
- online training while ANN is coupled to PDE.
- 1 data point per time step.

[2] Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1. 0). Geoscientific Model Development, 13(5), 2185-2196.

[3] Sahoo, D. et al, (2017). Online deep learning: Learning deep neural networks on the fly. arXiv preprint arXiv:1711.03705.

Online training

- online training while ANN is coupled to PDE.
- 1 data point per time step.
- First step: just do back propagation online:



- More sophisticated methods, See Rasp or Sahoo [2] [3].

[2] Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1. 0). Geoscientific Model Development, 13(5), 2185-2196.

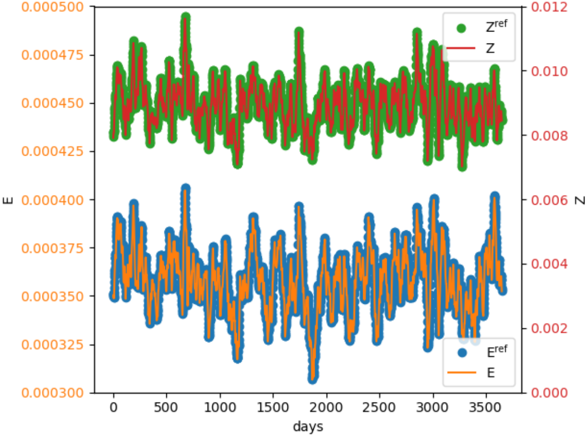
[3] Sahoo, D. et al, (2017). Online deep learning: Learning deep neural networks on the fly. arXiv preprint arXiv:1711.03705.

Initial results

- $M\Delta t$ = time interval between back propagation steps.

Initial results

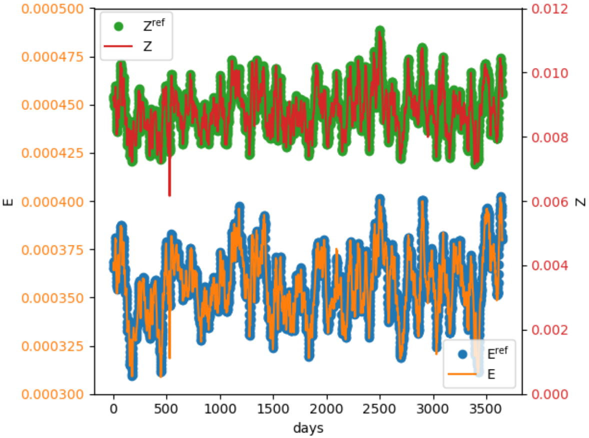
- $M\Delta t$ = time interval between back propagation steps.
- $M = 1$, <u>continual</u> online learning:



- <u>coupled</u> LR - ML model conserves HR energy and enstrophy.

Initial results

- $M = 20$:



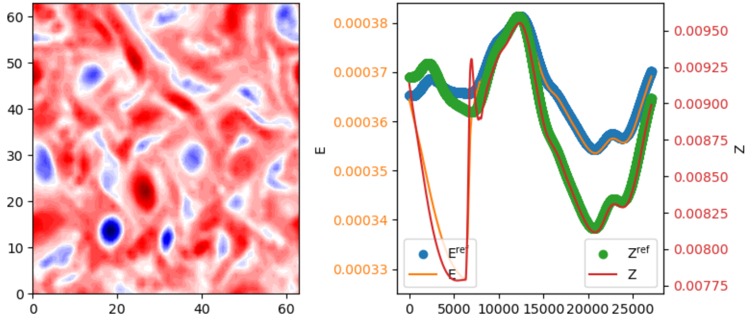
- coupled LR - ML model conserves HR energy and enstrophy.

Initial results

- Thus far, initial conditions were perfect: $q^{ref} = q$.
- Can it recover when $q^{ref} \neq q$?

Initial results

- Thus far, initial conditions were perfect: $q^{ref} = q$.
- Can it recover when $q^{ref} \neq q$?
- No SGS term before 10 days:



- coupled LR - ML model can quickly recover HR QoI.

Next steps

One of the next steps:

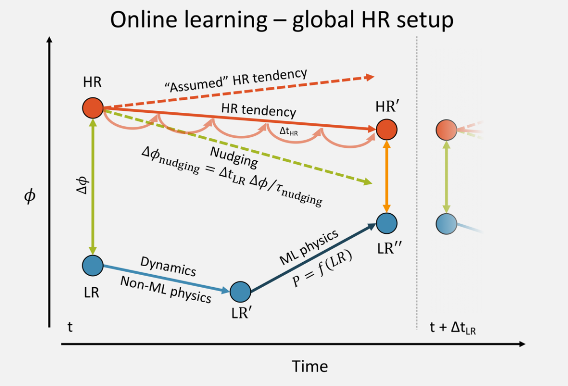
- When do we turn online training on / off? Is a UQ question.

Questions?

Edeling, W., & Crommelin, D. (2020). Reducing data-driven dynamical subgrid scale models by physical constraints. Computers & Fluids, 201, 104470.

Online training: Rasp (2020)[4]

Fig from Rasp 2020: HR state is nudged towards LR state.

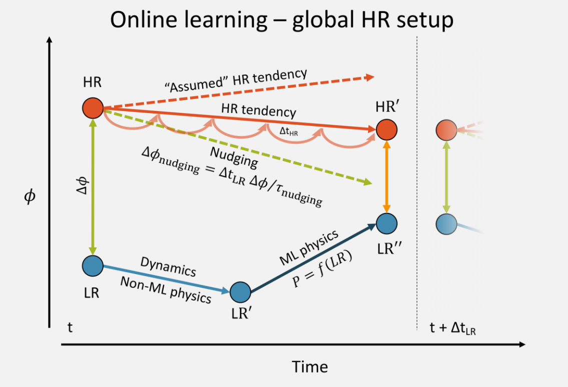


- Keeps HR - LR small, helps with online convergence.
- Allows LR to learn 'what HR would do under similar states'.
- Applied to Lorenz96.

[4] Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1. 0). Geoscientific Model Development, 13(5), 2185-2196.

Online training: Rasp (2020)[4]

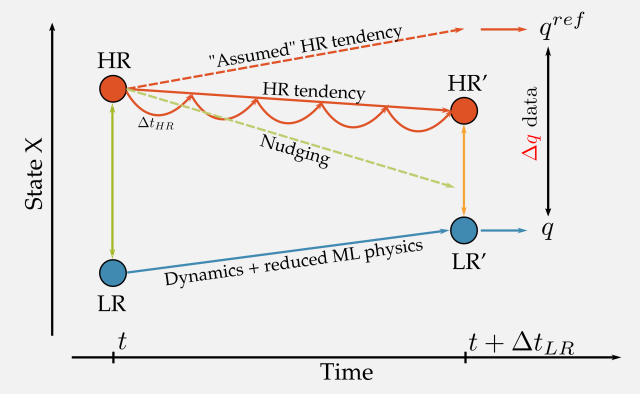
Fig from Rasp 2020: HR state is nudged towards LR state.



- However: nudging & ML correction is applied to entire state.
- Reduced ML: ML correction is only applied via τ.

[4] Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1. 0). Geoscientific Model Development, 13(5), 2185-2196.

Reduced online training



- Modified online (reduced) learning:
  - → Extract Δq data.
  - → Only 1 LR step.

Example results

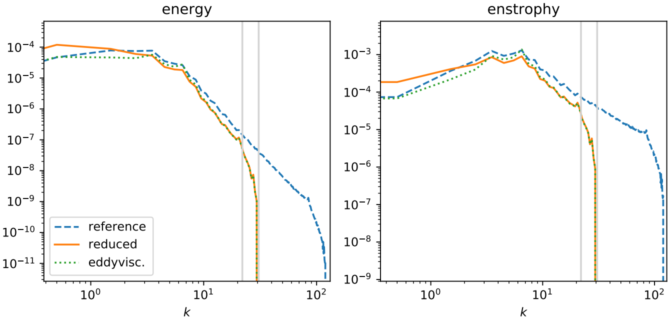
- Are the results spectrally accurate?
- Map 2D wave numbers $\mathbf{k} = (k_1, k_2)$ to 1D k:

$$k - \frac{1}{2} \leq \sqrt{k_1^2 + k_2^2} < k + \frac{1}{2}, \quad k = 0, 1, \cdots, \mathrm{ceil}\left(\sqrt{2}K\right)$$

Example results

- Are the results spectrally accurate?
- Map 2D wave numbers $\mathbf{k} = (k_1, k_2)$ to 1D k:

$$k - \frac{1}{2} \leq \sqrt{k_1^2 + k_2^2} < k + \frac{1}{2}, \quad k = 0, 1, \cdots, \mathrm{ceil}\left(\sqrt{2}K\right)$$



- <u>No</u> accurate spectra, only explicitly track overall E and Z.

Example results

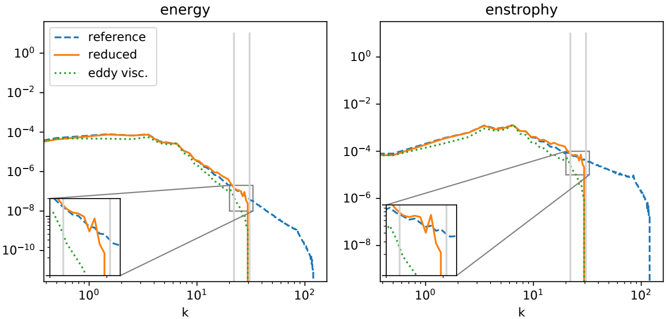
- However, scale-aware QoI are also possible
- Focus on a specific wave-number range via (spectral) filter $\mathcal{T}$:
  $\rightarrow$ zeros out all wave numbers $k < K$.

$$\tau_i \left(\frac{\partial f_i}{\partial \omega}, P_i \right) = \Delta q_i \rightarrow \tau_i \left(\mathcal{T} \frac{\partial f_i}{\partial \omega}, P_i \right) = \mathcal{T}(\Delta q_i)$$

Example results

- However, scale-aware QoI are also possible
- Focus on a specific wave-number range via (spectral) filter $\mathcal{T}$:
  $\rightarrow$ zeros out all wave numbers $k < K$.

$$\tau_i\left(\frac{\partial f_i}{\partial \omega}, P_i\right) = \Delta q_i \rightarrow \tau_i\left(\mathcal{T}\frac{\partial f_i}{\partial \omega}, P_i\right) = \mathcal{T}(\Delta q_i)$$



- Focus on wave numbers $k \in [21, 30]$.