Learning Stochastic Multiscale Models

Andrew F. Ilersich
andrew.ilersich@mail.utoronto.ca

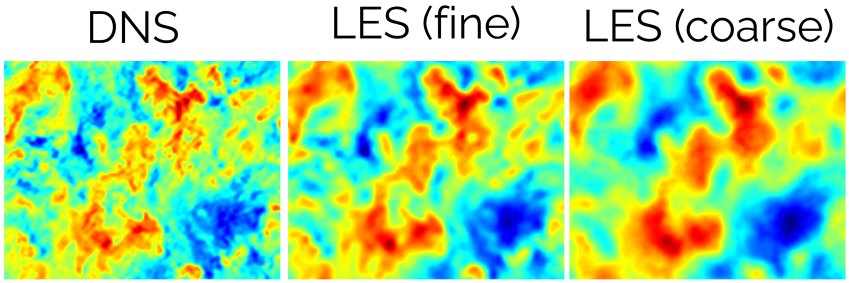
Prasanth B. Nair
prasanth.nair@utoronto.ca

NeurIPS 2025

Institute for Aerospace Studies
UNIVERSITY OF TORONTO

NEURAL INFORMATION PROCESSING SYSTEMS

Background



[1]

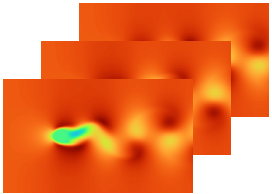
- Complex physical systems exhibit a vast range of scales
- Resolving the smallest scales is computationally infeasible, requiring approximation
- Closure models are a common approach that uses coarsened representations

[1] Tom-Robin Teschner. The introduction to LES I wish I had. CFD University. 2025.

Problem Setting

$u : \Omega \times [t_0, T] \to \mathbb{R}^{d_u}$
is a spatio-temporal field

Sample on fine spatial grid, over n_t time steps

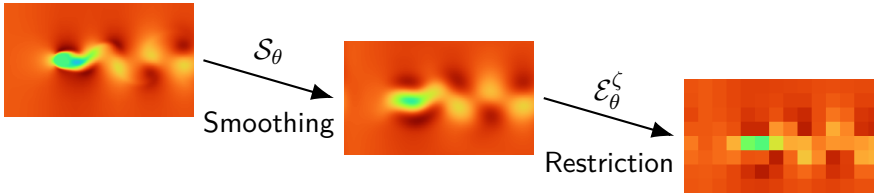


Training dataset $\mathcal{Y} = \{(t_i, y_i)\}_{i=1}^{n_t}$ where each snapshot $y_i \in \mathbb{R}^{n_y}$.

Goal

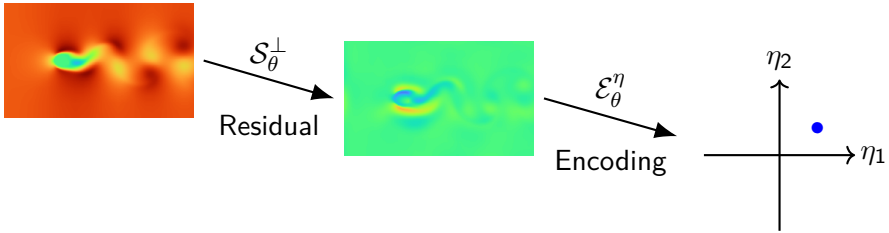
Given dataset $\mathcal{Y}$, learn a stochastic multiscale model that explicitly captures the coupled dynamics of **macroscale** and **microscale** features

Forming the Macroscale State with Probabilistic Encoder



Macroscale state $\zeta \sim p_\theta(\zeta \mid y)$, where $\mathbb{E}[\zeta \mid y] = \mathcal{E}_\theta^\zeta(\mathcal{S}_\theta(y))$

Forming the Microscale State with Probabilistic Encoder



Microscale state $\eta \sim p_\theta(\eta \,|\, y)$, where $\mathbb{E}[\eta \,|\, y] = \mathcal{E}_\theta^\eta(\mathcal{S}_\theta^\perp(y))$ and $\mathcal{S}_\theta^\perp(y) = y - \mathcal{S}_\theta(y)$

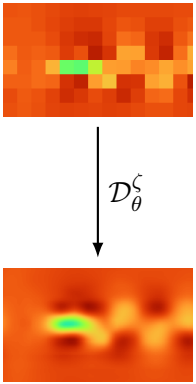
Stochastic Latent Dynamics

$$\mathrm{d}\begin{bmatrix} \zeta \\ \eta \end{bmatrix} = \begin{bmatrix} f_\theta(\zeta, \phi(\eta)) \\ g_\theta(\eta, \psi(\zeta)) \end{bmatrix} \mathrm{d}t + L_\theta(t)\mathrm{d}\beta$$

- Macroscale drift f_θ, microscale drift g_θ
- Dispersion matrix L_θ
- Wiener process β
- Inter-scale modulation terms ϕ and ψ

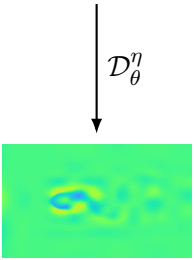
Reconstructing the Full State with Probabilistic Decoder

Prolong macroscale state mean



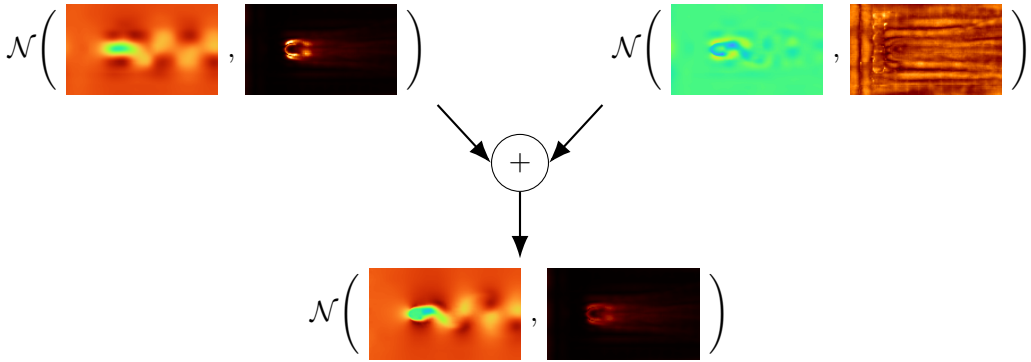
Decode microscale state mean

$\{\eta_1, \eta_2, \dots\}$



Reconstructing the Full State with Probabilistic Decoder

Combine large-scale and small-scale features



Multiscale Likelihood

Goal

Learn a multiscale representation with explicit scale separation

Multiscale Likelihood

Goal

Learn a multiscale representation with explicit scale separation

Standard likelihoods do not ensure scale separation

$$p_\theta(y \mid \zeta, \eta) = \mathcal{N}(y \mid \mathcal{D}_\theta^\zeta(\zeta) + \mathcal{D}_\theta^\eta(\eta), \Sigma_y) \quad \text{✗}$$

Multiscale Likelihood

Goal

Learn a multiscale representation with explicit scale separation

Standard likelihoods do not ensure scale separation

$$p_\theta(y \mid \zeta, \eta) = \mathcal{N}(y \mid \mathcal{D}_\theta^\zeta(\zeta) + \mathcal{D}_\theta^\eta(\eta), \Sigma_y) \quad \text{✗}$$

We formulate a multiscale likelihood based on Product of Experts (PoE) approach:

$$p_\theta(y \mid \zeta, \eta) = \frac{1}{Z(\zeta, \eta)} \times \underbrace{\mathcal{N}\left(y \;\middle|\; \mathcal{D}_\theta^\zeta(\zeta), \Sigma_\zeta\right)}_{\text{Macroscale expert}} \times \underbrace{\mathcal{N}\left(y - \mathcal{D}_\theta^\zeta(\zeta) \;\middle|\; \mathcal{D}_\theta^\eta(\eta), \Sigma_\eta\right)}_{\text{Microscale expert}}$$

The microscale state only captures what the macroscale cannot

Multiscale Likelihood

Goal

Learn a multiscale representation with explicit scale separation

$$\log p_\theta(y \mid \zeta, \eta) = \underbrace{-\frac{1}{2}\|y - \mathcal{D}_\theta^\zeta(\zeta)\|^2_{\left(\Sigma_\zeta^{-1}+\Sigma_\eta^{-1}\right)}}_{\text{Macroscale residual}} + \underbrace{\left(y - \mathcal{D}_\theta^\zeta(\zeta), \mathcal{D}_\theta^\eta(\eta)\right)_{\Sigma_\eta^{-1}}}_{\text{Microscale correlation}}$$

$$\underbrace{-\frac{1}{2}||\mathcal{D}_\theta^\eta(\eta)||^2_\Lambda}_{\text{Microscale norm}} + \underbrace{\frac{1}{2}\log\left|\Sigma_\zeta^{-1} + \Sigma_\eta^{-1}\right|}_{\text{Cov. determinant}} - \frac{n_y}{2}\log(2\pi)$$

where $\|v\|_A^2 = v^T A v$, $(u, v)_A = u^T A v$, and $\Lambda = \Sigma_\eta^{-1}(\Sigma_\zeta^{-1} + \Sigma_\eta^{-1})^{-1}\Sigma_\eta^{-1}$.

Numerical Results

Comparison of error statistics for each model on each test case, OA = our approach

	Error mean $\pm$ std. dev. on test set			
Model type	Wave 1D	KdV 1D	Cylinder 2D	SWE 2D
Coarse DNS	1.381 ± 0.382	0.523 ± 0.196	0.130 ± 0.040	0.861 ± 0.190
DMD	0.492 ± 0.066	0.163 ± 0.046	0.016 ± 0.006	0.531 ± 0.227
POD-SINDy	0.469 ± 0.054	0.057 ± 0.028	0.051 ± 0.024	0.535 ± 0.210
Macro-only	0.565 ± 0.116	0.324 ± 0.078	0.063 ± 0.005	0.426 ± 0.099
OA, $n_\eta = 5$	$\mathbf{0.079 \pm 0.031}$	$\mathbf{0.042 \pm 0.017}$	$\mathbf{0.009 \pm 0.001}$	$\mathbf{0.152 \pm 0.047}$

Explicitly modelling the microscale state improves error across all test cases

Visualization of Scale Separation

Learning Stochastic Multiscale Models

Code and data available:
`github.com/ailersic/multiscale-visde`