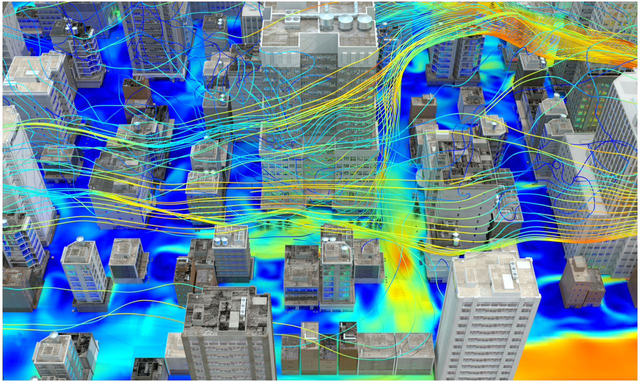
A probabilistic approach to subgrid-scale modeling

Jiaxi Zhao
joint with Sohei Arisaka and Qianxiao Li @ NUS

AI4X
July 9, 2025

Fluid simulations in real world

Full-resolution simulations are unaffordable for climate and urban environment modeling.



Can we design or learn better SGS models based on the fine-grid simulation data so that it can achieve accurate results even on coarse grid simulations?

A primer on SGS modeling

$$\begin{aligned} \frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j}(u_i u_j) &= -\frac{\partial p}{\partial x_i} + \nu \Delta u_i, \\ \frac{\partial u_i}{\partial x_i} &= 0. \end{aligned}$$

Applying a filter G to the equation, i.e. $\overline{u} = G * u$ (rough to smooth)

$$\begin{aligned} \frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}(\overline{u}_i \overline{u}_j) &= -\frac{\partial \overline{p}}{\partial x_i} + \nu \Delta \overline{u}_i - \frac{\partial \tau_{ij}}{\partial x_j}, \\ \frac{\partial \overline{u}_i}{\partial x_i} &= 0, \\ \overline{\mathbf{u}} \longrightarrow \tau_{ij} &= \overline{u_i u_j} - \overline{u}_i \overline{u}_j. \end{aligned}$$

Classical SGS stress models include the (dynamic) Smagorinsky model, and the Wall-Adapting Local Eddy-Viscosity model, and etc. They all based on physical intuition and empirical data.

Pipeline of Data-driven SGS modeling

1. Conduct fine-grid simulations and apply filter to obtain data pair of filtered-velocity and SGS stress.
2. Train a SGS model.
$$\min_{\phi} \|\tau - \phi(\overline{\mathbf{u}})\|^2 .$$
3. Deploy the model in the simulation.

$$\begin{aligned} \frac{\partial \overline{u}_i}{\partial t} + \frac{\partial}{\partial x_j}(\overline{u}_i \overline{u}_j) &= -\frac{\partial \overline{p}}{\partial x_i} + \nu \Delta \overline{u}_i - \frac{\partial \phi(\overline{u})_{ij}}{\partial x_j}, \\ \frac{\partial \overline{u}_i}{\partial x_i} &= 0. \end{aligned}$$

We care more about the performance of the SGS model in simulation (a-posteriori performance) than training loss (a-priori performance).

Existing modeling approaches

1. Rely on classical turbulence models, e.g. Smagorinsky model and use DNS data to fit the model parameters.

$$\tau = C_\theta(\overline{\mathbf{u}})||S||^2 S, \quad S = \nabla\overline{\mathbf{u}} + \nabla\overline{\mathbf{u}}^T.$$

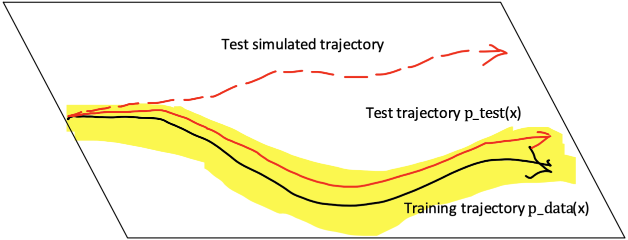
2. Choose the input features and fit an end-to-end mapping between them and the QoI.

$$\tau = \phi_\theta(\overline{\mathbf{u}}).$$

3. View the SGS stress modeling as a policy and solve it in the framework of reinforcement learning.

A-priori and a-posteriori discrepancy

The inconsistency between the a priori error and a posteriori error arises because the training algorithm does not take the solver dynamics into account.



p_data(x) ≠ p_test(x)

TABLE 3. Network and performance details

	Network inputs	Network outputs	*A priori* correlations	*A posteriori* simulations
NN-1	Local $\bar{S}_{ij}$	$\partial_j \tau_{ij}$	0.6	Stable; varying accuracy
NN-2	19-point stencil $\bar{S}_{ij}$	$\partial_j \tau_{ij}$	0.9	Unstable
NN-3	Local $\bar{S}_{ij}$	$L_i - D_i$ (Eq. 2.4)	0.7	Unstable

Visualizing SGS modeling: KS equation

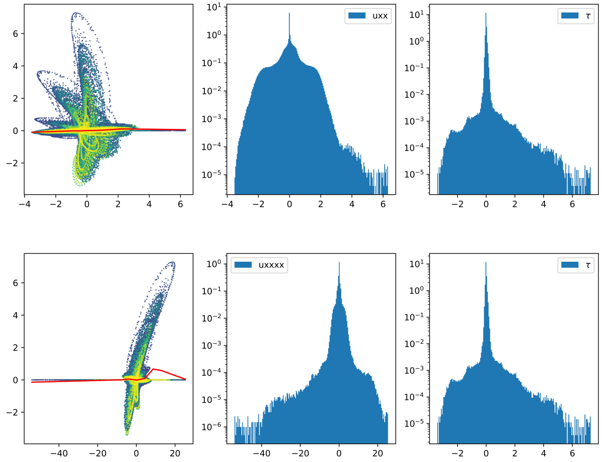
1D toy model: Kuramoto–Sivashinsky equation.

$$
\begin{aligned}
u_t &= -(c+u)u_x - uu_x - u_{xx} - \nu u_{xxxx}, \\
u(0,t) &= u(L,t) = 0, \\
u_x(0,t) &= u_x(L,t) = 0, \forall t.
\end{aligned}
$$

1. Fine-grid simulation: u_{1024};
2. Restriction operator: P to coarse grid;
3. Filtered variables: $u_{256} = Pu_{1024}$.
4. SGS stress:
$$\tau_{256} = \mathrm{P}(u_{1024}^2) - (\mathrm{P}\, u_{1024})^2.$$

Multivalue issue

Both $\mathbf{u}$ and τ are 1D, scatter to visualize them.



Same input can produce different outputs.

Solution: Probabilistic ansatz

Regression to generative modeling:

$$\tau = \phi_\theta(\mathbf{u}) \quad \to \quad \tau \sim p_\theta(\cdot|\mathbf{u}).$$

Change of the loss functions:

$$\min_\theta \sum_n \left\| \phi_\theta(\widetilde{\mathbf{u}}^{(n)}) - \tau^{(n)} \right\|^2, \quad \max_\theta \sum_{i=n}^{N} \log p_\theta(\tau^{(n)}|\mathbf{u}^{(n)}).$$

[2] Zhao, Jiaxi, Sohei Arisaka, and Qianxiao Li. "Generative subgrid-scale modeling." ICLR 2025 Workshop on Machine Learning Multiscale Processes.

Optimizing the model

Instead of using the MSE as the loss function, we optimize our model by maximize the likelihood over data:

$$\min_{\theta} \sum_{i=1}^{N} \frac{(y_i - \mu_\theta(x_i))^2}{2(\sigma_\theta(x_i))^2} + \log \sigma_\theta(x_i),$$

$$\min_{\theta} \sum_{i=1}^{N} -\log \left(\sum_{j=1}^{M} \frac{\text{softmax}(c_\theta^j(x_i))}{\sigma_\theta^j(x_i)} \exp\left[-\frac{(y_i - \mu_\theta^j(x_i))^2}{2(\sigma_\theta^j(x_i))^2} \right] \right).$$

Currently, we naively use the Adam solver to optimize this non-linear optimization problem.

Integrating in the simulation

How can we deploy this model in the simulation?

$$
\begin{aligned}
\text{Gaussian: } u_i &\Longrightarrow \mu_\theta(u_i), \sigma_\theta(u_i), z \sim N(0,1), \\
\tau_{ij} &= \mu_\theta(u_i) + \sigma_\theta(u_i) z, \\
\text{Gaussian mixture: } u_i &\Longrightarrow \mu_\theta^j(u_i), \sigma_\theta^j(u_i), z \sim N(0,1), j \sim [M], \\
\tau_i &= \mu_\theta^j(u_i) + \sigma_\theta^j(u_i) z,
\end{aligned}
$$

There is temporal and spatial consistency issue. Should we use the same latent variable z for all the grid points at all the time step or we should use different z_i for different u_i?

Experiments results: KS equation

$$\langle \overline{u} \rangle = \frac{1}{LT}\int_{[0,L]}\int_t^{t+T} u(x,t)dtdx,$$

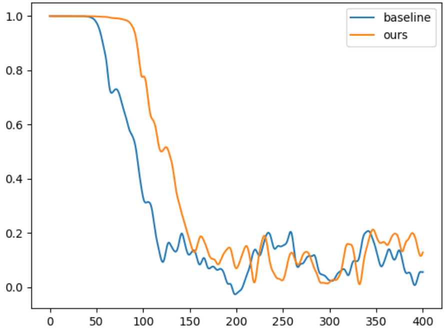
$$\left\langle \overline{u^2} \right\rangle = \frac{1}{LT}\int_{[0,L]}\int_t^{t+T} u^2(x,t)dtdx,$$

	baseline	regression	gaussian, fix	gaussian sample
A priori error	NA	0.976	-2.173	-2.173
$\int \lVert \mathbf{u} - \mathbf{u}_0 \rVert_2^2 \, dxdt$	**1.524**	2.036	1.720	1.597
$\langle \overline{u} \rangle - \langle \overline{u_0} \rangle$	9.901E-02	1.011E-01	3.870E-02	**3.214E-02**
$\left\langle \overline{u^2} \right\rangle - \left\langle \overline{u_0^2} \right\rangle$	-4.326E-01	-4.241E-01	-1.895E-01	**-6.577E-02**
A priori error	NA	0.987	-2.583	-2.583
$\int \lVert \mathbf{u} - \mathbf{u}_0 \rVert_2^2 \, dxdt$	**1.524**	1.817	1.898	1.889
$\langle \overline{u} \rangle - \langle \overline{u_0} \rangle$	9.901E-02	8.012E-02	**-1.242E-02**	-1.543E-02
$\left\langle \overline{u^2} \right\rangle - \left\langle \overline{u_0^2} \right\rangle$	-4.326E-01	-3.466E-01	**-1.018E-01**	-1.296E-01

First stage NS equation results

$$\mathrm{corr}(t) = \frac{\langle u^{\mathrm{c}}(\cdot, t), \overline{u^{\mathrm{f}}}(\cdot, t) \rangle}{\| u^{\mathrm{c}}(\cdot, t) \|_2 \left\| \overline{u^{\mathrm{f}}}(\cdot, t) \right\|_2},$$

Our method can achieve a better temporal correlation with the fine-grid simulations comparing to the regression-based methods.



Future work

1. Implement more expressive generative models for SGS modeling and investigate their performance.
2. Deploy to practical problems: Subgrid-scale modeling in large eddy simulation of the urban environment.
3. Theoretical understanding of the difference between regression-based and generative-based SGS modeling, especially in different application scenarios such as CFD and MD.

References and advertisement

1. Zhao, Jiaxi, and Qianxiao Li. "Mitigating Distribution Shift in Machine Learning–Augmented Hybrid Simulation." SIAM Journal on Scientific Computing 47.2 (2025): C475-C500.
2. Zhao, Jiaxi, Sohei Arisaka, and Qianxiao Li. "Generative subgrid-scale modeling." ICLR 2025 Workshop on Machine Learning Multiscale Processes. 2025.
3. Code repository: `https://github.com/jiaxi98/ml4dynamics`.

I am looking for both industrial and academic positions starting from this year, collaborations and discussions are also warmly welcome!

Email: jiaxi.zhao@nus.edu.sg; WeChat:

KS Statistics