Reduced-order Models and Data-driven Closure Strategies for Turbulent Systems

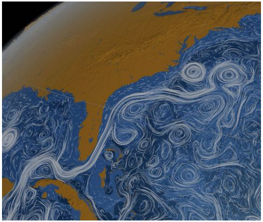
Di Qi
with Andrew Majda (NYU)
and John Harlim (Penn State), Nan Chen (UW-Madison)

Purdue University

Mathematical Approaches of Atmospheric Constitutes
Data Assimilation & Inverse Modeling, BIRS
March 21, 2023

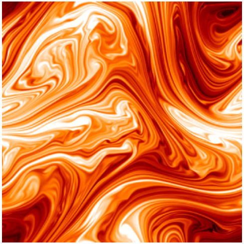
Introduction: complex turbulent systems

Turbulent dynamical systems are characterized by *a large dimensional phase space* and *high degrees of internal instability* (e.g., geoscience and plasma physics)



Challenges:

- *multiscale* with *strong nonlinear interactions*
- nonlinear interactions with a *non-Gaussian* equilibrium state
- understand and predict *extreme events*



Central math/science issues:

- quantifying uncertainty and model errors
- capture statistical variability to *general initial* and *external perturbations*
- learning extreme dynamics from data

[1] *E & Enguist, 2003; Majda & Wang, 2008; Needlin et al, 2011; Lucarini et al, 2020*

Outline

1. A reduced-order statistical model for general turbulent systems

2. Machine learning-based statistical closure model

3. Data-driven conditional Gaussian forecast and data assimilation

General framework for turbulent systems

The system setup will be a finite-dimensional state $\mathbf{u}(t;\omega) \in \mathbb{R}^N$ subject to linear dynamics and an energy preserving nonlinear part

$$\frac{d\mathbf{u}}{dt} = \mathcal{F}[\mathbf{u}(t;\omega);\omega] = (L + D)\mathbf{u} + B(\mathbf{u},\mathbf{u}) + \mathbf{F}(t) + \boldsymbol{\sigma}(t)\dot{W}(t;\omega) \quad (1)$$

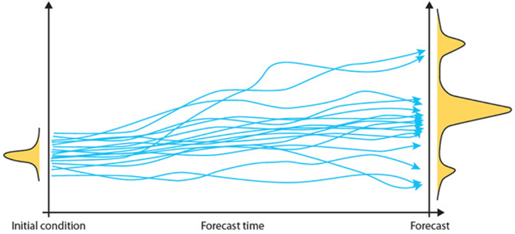
- skew-symmetric $L^* = -L$ (e.g. rotation, dispersion etc.)
- negative definite $D \leq 0$ (surface drag, viscosity, dissipations etc.)
- external forcing: deterministic $\mathbf{F}(t)$ (solar force, wind stress ...)
- unresolved effects: white noise $\boldsymbol{\sigma}(t)\dot{W}(t;\omega)$
- energy-conserving quadratic form: $\mathbf{u} \cdot B(\mathbf{u},\mathbf{u}) \equiv 0$

Statistical ensemble forecast

- A probabilistic forecast of the model states is needed for tracking the evolution of the PDFs
- Curse-of-dimensionality occurs in MC-type approaches, especially with non-Gaussian higher-order statistics
- In practice, ensemble forecast via data assimilation is essential, especially in the situation with partial observations

Important tasks:

- developing *statistical reduced order models*
- efficient *ensemble forecast* for PDFs



Exact statistical moment equations

$$\frac{d\mathbf{u}}{dt} = (L + D)\,\mathbf{u} + B(\mathbf{u}, \mathbf{u}) + \mathbf{F}(t) + \boldsymbol{\sigma}(t)\,\dot{W}(t;\omega)$$

Statistical dynamical equations for the mean and covariance

$$\mathbf{u}(t) = \bar{\mathbf{u}}(t) + \sum Z_i(t;\omega)\,\boldsymbol{v}_i : \quad \bar{\mathbf{u}} = \langle u \rangle_p \quad R_{ij} = \langle Z_i Z_j^* \rangle_p :$$

$$\begin{aligned}
\frac{d\bar{\mathbf{u}}}{dt} &= (L + D)\,\bar{\mathbf{u}} + B(\bar{\mathbf{u}}, \bar{\mathbf{u}}) + R_{ij} B(\boldsymbol{v}_i, \boldsymbol{v}_j) + \mathbf{F}(t), \\
\frac{dR}{dt} &= L_v(\bar{\mathbf{u}})\,R + R L_v^*(\bar{\mathbf{u}}) + Q_F + \sum_k \boldsymbol{v}_i^* \sigma_k^* \cdot \sigma_k \boldsymbol{v}_j. \\
\frac{dZ_k}{dt} &= \sum_m L_{v,km}(\bar{u})\,Z_m + \sigma(t)\,\dot{W}(t;\omega) \cdot v_k \\
&\quad + \sum_{m,n} (Z_m Z_n - R_{mn})\,B(v_m, v_n) \cdot v_k.
\end{aligned} \tag{2}$$

Exact statistical moment equations

Statistical dynamical equations for the mean and covariance

$$\mathbf{u}(t) = \bar{\mathbf{u}}(t) + \sum Z_i(t;\omega)\, \boldsymbol{v}_i : \quad \bar{\mathbf{u}} = \langle u \rangle_p \quad R_{ij} = \langle Z_i Z_j^* \rangle_p$$

$$\begin{aligned} \frac{d\bar{\mathbf{u}}}{dt} &= (L+D)\,\bar{\mathbf{u}} + B(\bar{\mathbf{u}},\bar{\mathbf{u}}) + R_{ij} B(\boldsymbol{v}_i,\boldsymbol{v}_j) + \mathbf{F}(t), \\ \frac{dR}{dt} &= L_v(\bar{\mathbf{u}})\, R + R L_v^*(\bar{\mathbf{u}}) + Q_F + \sum_k \boldsymbol{v}_i^* \sigma_k^* \cdot \sigma_k \boldsymbol{v}_j. \end{aligned} \tag{3}$$

- the linear operator L_v expressing energy transfers between the *mean field and the stochastic modes* (B), *dissipation* (D), and *non-normal dynamics* (L)

$$\{L_v(\bar{\mathbf{u}})\}_{ij} = \left[(L+D)\,\boldsymbol{v}_j + B(\bar{\mathbf{u}},\boldsymbol{v}_j) + B(\boldsymbol{v}_j,\bar{\mathbf{u}})\right] \cdot \boldsymbol{v}_i.$$

- the nonlinear flux operator Q_F for third-order moments expressing the energy flux due to non-linear terms

$$Q_{F,ij} = \sum_{m,n} \langle Z_m Z_n Z_j \rangle B(\boldsymbol{v}_m,\boldsymbol{v}_n) \cdot \boldsymbol{v}_i + \langle Z_m Z_n Z_i \rangle B(\boldsymbol{v}_m,\boldsymbol{v}_n) \cdot \boldsymbol{v}_j.$$

Ideas for Reduced-Order Statistical Energy Closure

The reduced-order approximation $u_M \in \mathbb{R}^M,\ M \ll N$

$$\begin{aligned}\frac{d\bar{u}_M}{dt} &= (L+D)\,\bar{u}_M + B(\bar{u}_M,\bar{u}_M) + R_{M,ij}B(v_i,v_j) + F,\\ \frac{dR_M}{dt} &= L_v(\bar{u}_M)\,R_M + R_M L_v^*(\bar{u}_M) + Q_F^M + Q_\sigma.\end{aligned}$$

A new systematic approach for the nonlinear flux Q_F^M combining both the *detailed model energy mechanism* and *control over model sensitivity*

$$\boxed{Q_F^M = Q_F^{M,-} + Q_F^{M,+}}$$

- Model fidelity guarantees convergence to the unperturbed equilibrium
- Model sensitivity quantifies responses to general external perturbations
  - *response operator* independent of specific perturbations[1]
  - *relative entropy* as the distance between two probability densities[2]

[1] *Ruelle, Nonlinear, 2009; Leith, JAS, 1975*

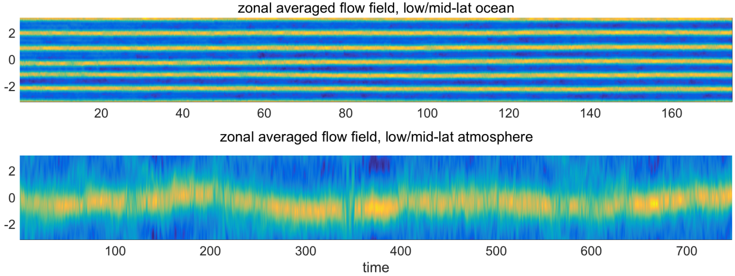
[2] *Abramov, Majda, Kleeman, JAS, 2005*

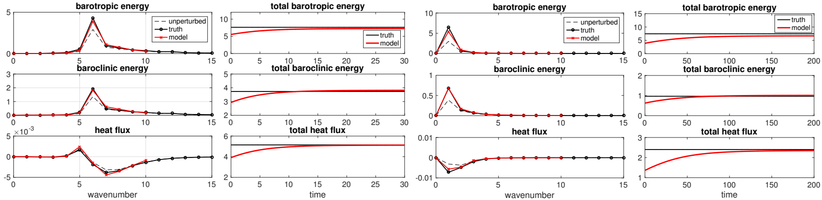
[3] *Sapsis & Majda, PNAS, 2013; Qi & Majda, SIAM Review, 2018*

Flow in low-latitude regimes with zonal jets

Deterministic forcing through a perturbation in the background shear δU

- True statistics from a DNS code with $256 \times 256 \times 2$ grid points
- In the reduced-order model, only modes $|k| \leq 10$ are resolved, which is about 0.15% of the full model resolution





Machine learning strategies for higher order statistics

Accurate prediction of key statistics in turbulent systems remains a challenging problem

- *non-Gaussian* statistical states
- interaction among a wide spectrum of scales
- curse of dimensionality

Machine learning strategies have been extensively applied to problems involving big data

- compositions of simple functions
- successful for learning dynamics
- data-driven predictions of turbulent systems

A machine learning strategy for high order responses in statistical closure models

- neural network is used to learn the nonlinear dynamics directly from data
- unresolved nonlinear flux in different scales are modeled automatically
- the method requires robust performance with internal instability

[1] *Ma, Wang, E, 2018; Levine, Stuart, 2021*

Non-Markovian model with neural network

full moment equations

$$\dot{\bar{u}} = (\mathcal{L} + \mathcal{D})\,\bar{u} + B(\bar{u}, \bar{u}) + \phi + F$$
$$\dot{R} = L_v(\bar{u})\,R + RL_v^*(\bar{u}) + \theta$$
$$\phi = R_{ij}B(v_i, v_j), \quad \theta_{ij} = \langle Z_m Z_n Z_j \rangle B(v_m, v_n) \cdot v_i + \text{c.c.}$$

Using a hidden non-Markovian model that maps the delay coordinates of variables $\bar{u}, R$ to nonlinear coupling ϕ, θ in a low-dimensional subspace

discrete low-order closure

$$\bar{u}_{i+1} = \mathcal{F}_1(\bar{u}_i, R_i, F_{i+1}, \phi_i), \ R_{i+1} = \mathcal{F}_2(\bar{u}_i, R_i, \theta_i)$$
$$\phi_{i+1} = \mathcal{G}_\phi(\bar{u}_{i-m:i}, R_{i-m:i}, \phi_{i-m:i})$$
$$\theta_{i+1} = \mathcal{G}_\theta(\bar{u}_{i-m:i}, R_{i-m:i}, \theta_{i-m:i})$$

Challenge: how to effectively learn the (nonlinear) structures in $\mathcal{G}_\phi, \mathcal{G}_\theta$ from data?

Connection to Mori-Zwanzig formalism

Suppose the memory length $m\,(< i)$ the delay embedding theorem holds

- The full dynamics as coupled resolved-unresolved processes

$$
\begin{aligned}
u_{i+1} &= \mathcal{F}\left(u_i, \theta_i\right), \\
\theta_{i+1} &= \mathcal{G}\left(u_i, \theta_i\right) \coloneqq \mathbb{E}\left(\Theta_{i+1} \mid u_{i-m:i}, \theta_{i-m:i}\right).
\end{aligned}
$$

- The approximation model to delay embedded map

$$
\hat{u}_{i+1} = \mathcal{F}\left(\hat{u}_i, \hat{\theta}_i\right), \quad \hat{\theta}_{i+1} = \mathbb{E}^{\epsilon}\left(\Theta_{i+1} \mid u_{i-m:i}, \theta_{i-m:i}\right) + \hat{\xi}_{i+1},
$$

where $\mathbb{E}^{\epsilon}$ is the estimator with variance of order ϵ^2 and $\hat{\xi}$ a noise.

- For the same initial condition, there is the error estimate

$$
\mathbb{E}\left(\max_{i\in[0,\cdots,T]} \left|\hat{u}_i - u_i\right|\right) = O\left(a^T \epsilon\right),
$$

where $a > 1$ is a constant that is independent of T and ϵ.

Training and prediction for turbulent systems

Basic idea:

- *training stage*: small training data from unperturbed equilibrium: i) constant initial value; ii) constant external forcing
- *prediction stage*: different initial & inhomogeneous perturbations beyond the training dataset among different perturbation scenarios

Neural network should maintain numerically stable to cope with the inherent instability persistent in the turbulent model.

Question: what is the skill in the optimized neural network to predict the highly nonlinear statistical responses using limited data set?

[1] *Qi, Harlim, Philos. Trans. R. Soc. A, 2021*

Architecture of the deep neural network

Long-Short-Term-Memory (LSTM) as a special recurrent neural network

- LSTM chain connected by m sequential cells

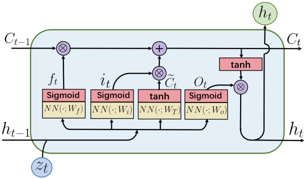
$$h_m = \mathrm{Lc}^{(m)}\{h_0; x_{i-m+1}, \cdots, x_i\} \equiv \mathrm{Lc}(x_i) \circ \cdots \circ \mathrm{Lc}(x_{i-m+1})(h_0)$$

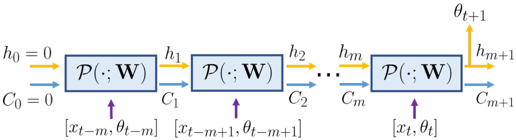
- a fully connected final layer

$$\hat{y}_{i+1} = A h_m + b$$

- loss function from relative entropy

$$\mathcal{P}\left(\pi, \pi^M\right) = \int \pi \ln\left(\pi/\pi^M\right) = \frac{1}{2}\left(\bar{u}_t - \bar{u}_m\right)^T R_m^{-1}\left(\bar{u}_t - \bar{u}_m\right) + \frac{1}{2}\left[\mathrm{tr}\left(R_t R_m^{-1}\right) - \log\det\left(R_t R_m^{-1}\right) - N\right].$$

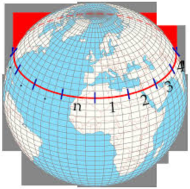


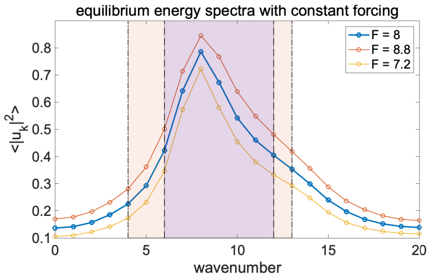


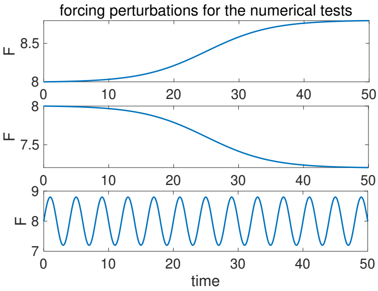
Example: Lorenz 96 system

The Lorenz 96 model mimics the large-scale behavior around a mid-latitude atmosphere circle

$$\frac{du_j}{dt} = u_{j-1}\left(u_{j+1} - u_{j-2}\right) - du_j + F.$$

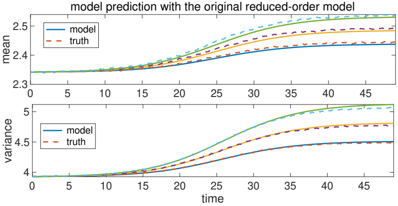


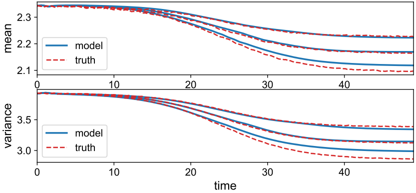


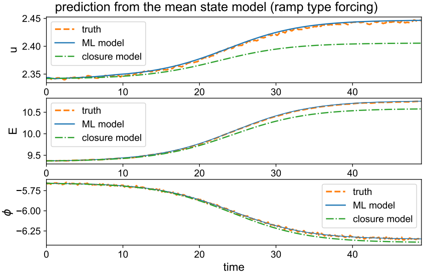


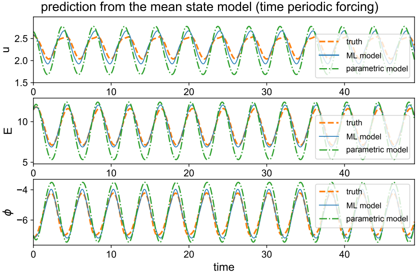
Model performance in the L-96 system

Machine learning prediction with *a much larger integration step* $\Delta t = 10 dt$









Additional issues with inhomogeneous statistics

Limitation in using the data-driven statistical models involving highly turbulent signals:

- Additional constraints in statistical moments:
  - positive-definite covariance, inhomogeneous higher moments
- Strong inherent instability among a wide range of fluctuation modes:
  - amplification of small errors, numerical instability
- Efficient ensemble simulations for data assimilation and filtering

Idea: modeling uncertainty from a hybrid statistical-stochastic formulation[3]

- Key leading order moments in explicit statistical dynamics
- High order fluctuation feedbacks from efficient stochastic closure model

[3] *Qi & Harlim, JCP, 2023*

A coupled statistical-stochastic model

The statistical-stochastic model can naturally estimate inhomogeneous statistics and positive-definite covariance estimation

$$\frac{d\bar{u}}{dt} = (L + D)\,\bar{u} + B(\bar{u},\bar{u}) + \sum_{i,j}\left(\frac{1}{M-1}\sum_{i=1}^{M} Z_k^{(i)} Z_l^{(i)*}\right) B(e_i, e_j) + F,$$

$$\frac{dZ_i^{(i)}}{dt} = \sum_j L_{ij}(\bar{u})\, Z_j + \sum_{m,n}\gamma_{imn}\left(Z_m^{(i)} Z_n^{(i)*}\right) + \sigma(t)\,\dot{W}^{(i)}(t;\omega)\cdot e_i.$$

The fluctuation equation has the equivalent covariance dynamics

$$\frac{dR}{dt} = L(\bar{u})\, R + RL^*(\bar{u}) + Q_F + Q_\sigma,$$

with mean-fluctuation decomposition and ensemble approximation

$$u = \bar{u} + u'(t;\omega) = \bar{u} + \sum_{i=1}^{N} Z_i(t;\omega)\, e_i, \quad R = \langle ZZ^*\rangle \sim \frac{1}{M-1}\sum_{i=1}^{M} Z^{(i)} Z^{(i)*}.$$

Overfitting in direct training of stochastic processes

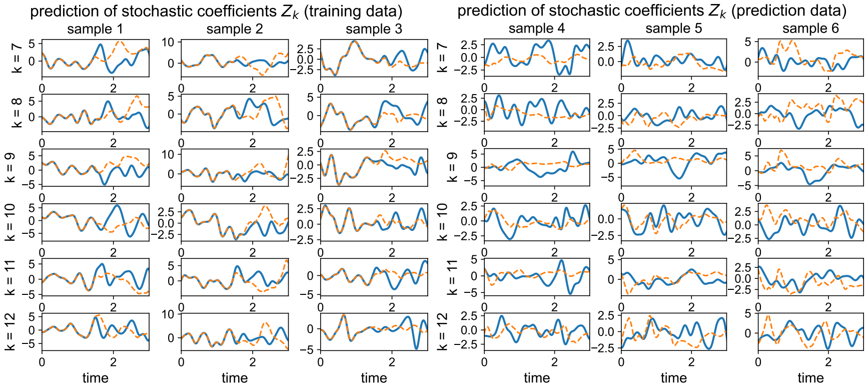


Figure: Trajectory prediction of stochastic coefficients Z_k in leading modes

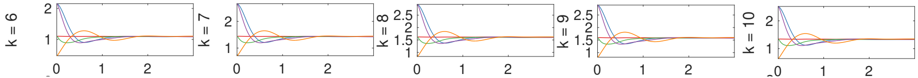


Figure: Lyapunov exponents computed for each spectral mode k of the L96 model

A stabilized reduced order closure

A low-dimensional representation in mean and fluctuation

$$u^M = \bar{u}^M + \sum_{i\in\mathcal{I}} Z_i e_i, \quad |\mathcal{I}| \ll N.$$

- *mean equation* in low-dimensional resolved subspace

$$\frac{d\bar{u}^M}{dt} = (L + D)\,\bar{u}^M + \sum_{i,j\in\mathcal{I}} R_{ij}^M B\,(e_i, e_j) + F \ + \boldsymbol{\Theta}^m$$

- a reduced order *fluctuation equation*

$$\frac{dZ^M}{dt} = L\left(\bar{u}^M\right) Z^M + \sigma\,(t)\,\dot{W}\,(t;\omega)\cdot e_i \ + \boldsymbol{\Theta}^\nu$$

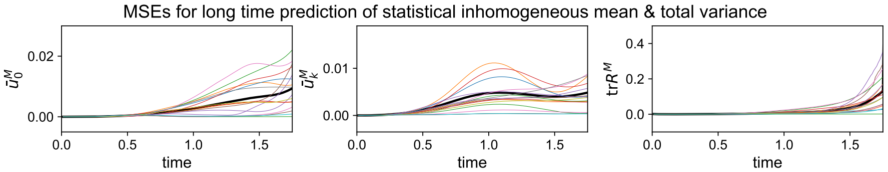
Decomposition of effective damping and noise

$$Q_{F,k} \approx \boldsymbol{\Theta}^\nu = -D^M Z^M + \Sigma^M \dot{W}.$$

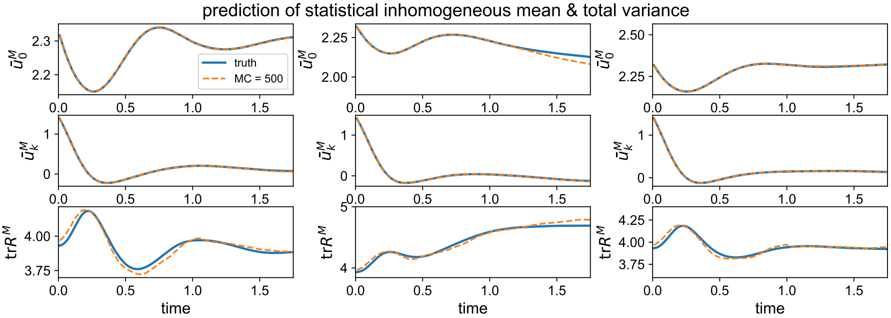
Equivalently, this gives

$$Q_F \approx -D^M R^M + R^M D^{M*} + \Sigma\Sigma^*.$$

Training and prediction results



(a) prediction MSEs in different test cases



(b) prediction of the inhomogeneous statistics in 3 cases

Physics-informed data-driven conditional Gaussian algorithm with partial observation

Goals:

- Efficiently and accurately forecasting *key non-Gaussian PDF* for a wide class of high-dimensional complex systems using only *a small number of ensemble samples*.
- Providing a systematic framework of developing *hybrid dynamical-statistical reduced order models* for complex systems with very large dimensions when the primary interest lies in the statistical forecast of certain large-scale modes.

Ensemble forecast with conditional Gaussian Model

- Ensemble prediction with the conditional Gaussian framework[1]

$$\begin{aligned}\frac{dX}{dt} &= \left[A_0\left(X,t\right) + A_1\left(X,t\right)Y\right] + B_1\left(X,t\right)\dot{W}_1,\\ \frac{dY}{dt} &= \left[a_0\left(X,t\right) + a_1\left(X,t\right)Y\right] + b_1\left(X,t\right)\dot{W}_2,\end{aligned}$$

with

$$p\left(X,Y\right) \sim \frac{1}{J}\sum_{j=1}^{J} p\left(X^{(j)}\right)\mathcal{N}\left(\mu_Y\left(X^{(j)}\right), R_Y\left(X^{(j)}\right)\right).$$

- Learning unresolved processes in the conditional Gaussian equations[2]

$$\begin{aligned}\frac{d\mu}{dt} &= \left(a_0 + a_1\mu\right) + \left(RA_1^*\right)\left(B_1B_1^*\right)^{-1}\left(\dot{X} - \left(A_0 + A_1\mu\right)\right),\\ \frac{dR}{dt} &= a_1R + Aa_1^* + b_2b_2^* - \left(RA_1^*\right)\left(B_1B_1^*\right)^{-1}\left(A_1R\right).\end{aligned}$$

[1] *Chen & Majda, 2017*

[2] *Chen & Qi, 2023*

Ensemble forecast with conditional Gaussian Model

- Ensemble prediction with the conditional Gaussian framework[1]

$$\frac{dX}{dt} = \left[A_0(X,t) + A_1(X,t)Y\right] + B_1(X,t)\dot{W}_1,$$
$$\frac{dY}{dt} = \left[a_0(X,t) + a_1(X,t)Y\right] + b_1(X,t)\dot{W}_2,$$

with

$$p(X,Y) \sim \frac{1}{J}\sum_{j=1}^{J} p\left(X^{(j)}\right)\mathcal{N}\left(\mu_Y\left(X^{(j)}\right), R_Y\left(X^{(j)}\right)\right).$$

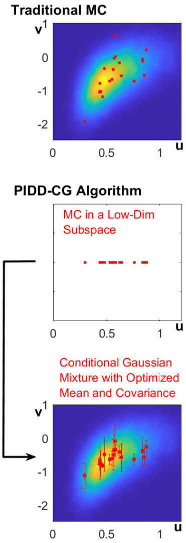
- Learning unresolved processes in the conditional Gaussian equations[2]

$$\frac{d\mu}{dt} = \left(a_0 + a_1\mu\right) + \mathcal{F}_Y\left(B_1B_1^*\right)^{-1}\mathcal{G}_Y,$$
$$\frac{dR}{dt} = a_1R + Aa_1^* + b_2b_2^* - \mathcal{F}_Y\left(B_1B_1^*\right)^{-1}\mathcal{F}_Y.$$

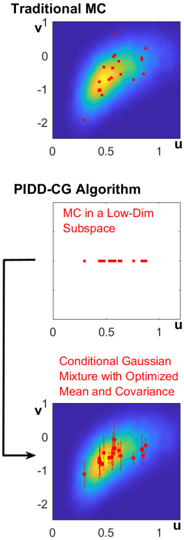
[1] *Chen & Majda, 2017*

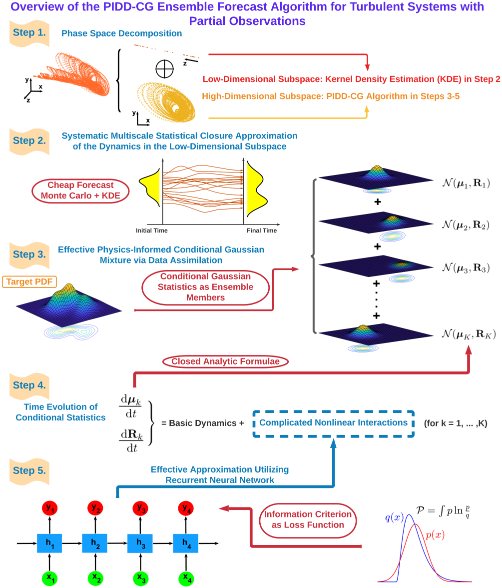
[2] *Chen & Qi, 2023*

General ideas



General ideas

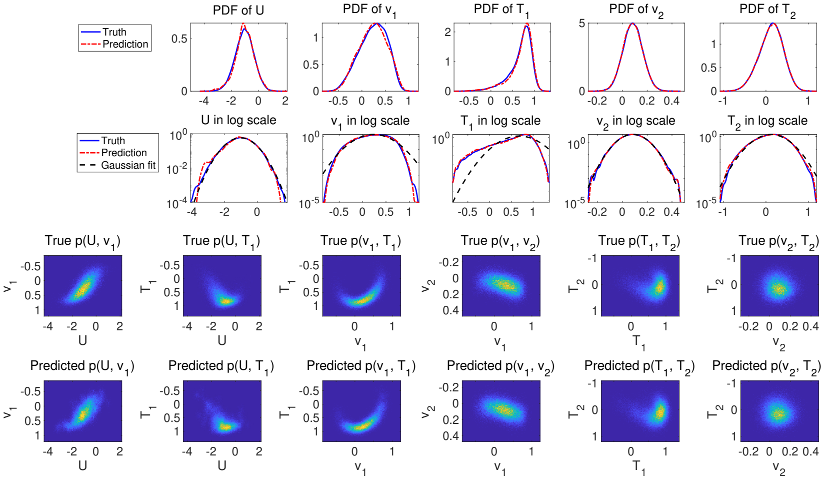




Prediction of marginal PDFs in turbulent transport

$$\frac{\partial q}{\partial t} + u \cdot \nabla q = \mathcal{D}(\Delta) q + F, \quad \frac{\partial T}{\partial t} + u \cdot \nabla T = -d_T T + \kappa \Delta T.$$

- Direct MC simulation: sample $N = 5 \times 10^4$, time step $\Delta t = 1 \times 10^{-3}$;
- Data-driven CG model: sample $N = 100$, time step $\Delta t = 0.01$.



Summary

- Prediction for higher order statistics becomes an important issue in strongly non-Gaussian regimes.
- Data-driven methods provide useful tool to effectively improve prediction skill and learn unresolved turbulent structures.
- The framework is also useful for stochastic modeling strategies combining ideas in statistical closure model, data assimilation, and conditional statistics ensemble.

Reference:

- Qi & Majda, *Using machine learning to predict extreme events in complex systems*, PNAS, 2020.
- Qi & Harlim, *Machine learning-based statistical closure of turbulent dynamical systems*, Philos. Trans. R. Soc. A, 2022.
- Qi & Harlim, *A data-driven statistical-stochastic model for effective ensemble forecast of complex turbulent systems*, JCP, 2023.
- Chen & Qi, *A physics-informed data-driven algorithm for ensemble forecast of complex turbulent systems*, preprint, 2023.