Paper Review of Data-assisted reduced-order modeling of extreme events in complex dynamical systems

Zhong Yi Wan, Pantelis Vlachas, Petros Koumoutsakos, Themistoklis Sapsis

Department of Mechanical Engineering, MIT
Chair of Computational Science, ETH

January 24, 2020

Table of Contents

Problem Setup

- $\frac{du}{dt} = F(u) = Lu + h(u)$, $u(t) \in \mathbb{R}^n$, where n is a large number.
- We are specifically interested in systems whose dynamics results in a intrinsically low dimensional globally attracting manifold.

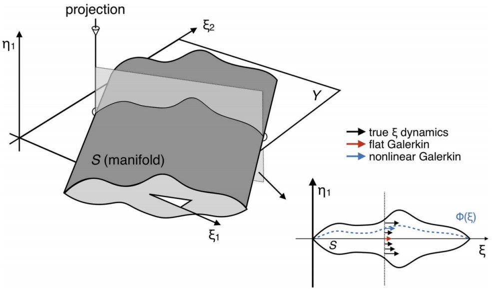


Figure: Example of a Globally Attracting Manifold

Problem Setup

Goal:

- Dimensionality Reduction of the ODE
- Derive an ODE in a low dimensional subspace, which preserves as much dynamical property of original ODE as possible.

Ansatz:

$$u = \mathbf{Y}\xi + \mathbf{Z}\eta + b \tag{1}$$

where the columns of matrix $\mathbf{Y} = [y_1, ..., y_m]$ form an orthonormal basis of Y, an m-dimensional subspace of $\mathbb{R}^d$, and the columns of $\mathbf{Z} = [z_1, ..., z_{d-m}]$ make up an orthonormal basis for the orthogonal complement $Z = \mathbb{R}^d \setminus Y$; ξ and η are the projection coordinates associated with $\mathbf{Y}$ and $\mathbf{Z}$.

Problem Setup

How to get **Y** and **Z**?

- Randomly sample 10000 points on the attractor, denoted as **X**.
- Do PCA on these datapoints:
  Let $C = X^T X$, then $\mathbf{Y} = [y_1, ..., y_m]$, where $[y_1, ..., y_m]$ are the eigenvectors of C corresponding to the m largest eigenvalues of C.
  $\mathbf{Z} = [z_1, ..., z_{d-m}]$, where $[z_, ..., z_{d-m}]$ are the rest of eigenvectors of C.

Back to equation (1), we know Y, Z and the ODE for u, can we have the ODE for ξ?

Table of Contents

Flat Galerkin Method

Plug (1) back to the ODE, we get

$$\frac{d\xi}{dt} = Y^T LY\xi + Y^T LZ\eta + Y^T h(Y\xi + Z\eta + b) + Y^T Lb \tag{2}$$

If $|\eta| << |\xi|$, then we may assume $\eta = 0$, leading to a m-dimensional system:

$$\frac{d\xi}{dt} = Y^T LY\xi + Y^T h(Y\xi + b) + Y^T Lb = F_\xi(\xi) \tag{3}$$

This is called the Flat Galerkin Method.

Nonlinear Galerkin Projection

Problems:

- $\eta = 0$ may be too strict.
- Z is derived merely based on statistical properties of the manifold without addressing the dynamics. This implies that even if η has small magnitude on average it may play a big role in the dynamics of the space.

Existing solution: Let $\eta = \Phi(\xi)$, then

$$\frac{d\xi}{dt} = Y^T L Y \xi + Y^T L Z \Phi(\xi) + Y^T h(Y\xi + Z\Phi(\xi) + b) + Y^T L b \tag{4}$$

This is called the Nonlinear Galerkin Projection Method.

Nonlinear Galerkin Projection

Problems:

- Φ has to be found empirically.
- Φ may not be well-defined!

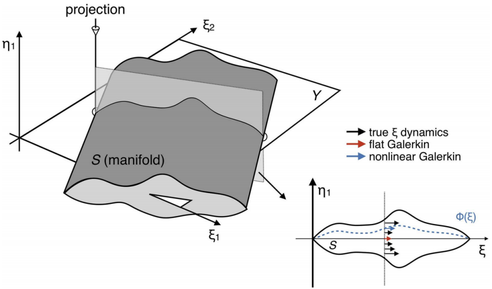


Figure: Example of a Globally Attracting Manifold

Data-driven Method

Proposed Data-driven solution:

$$\frac{d\xi}{dt} = Y^T L Y \xi + Y^T L Z \eta + Y^T h(Y\xi + Z\eta + b) + Y^T L b \tag{5}$$

rewritten as

$$\frac{d\xi}{dt} = F_\xi(\xi) + G(\xi, \eta) \tag{6}$$

Try to learn G from the data.
Assumption: $\Psi(t) = G(\xi(t), \eta(t)) \approx \hat{G}(\xi(t), \xi(t-\tau), \xi(t-2\tau), ...)$
What is $\hat{G}$? LSTM...

LSTM

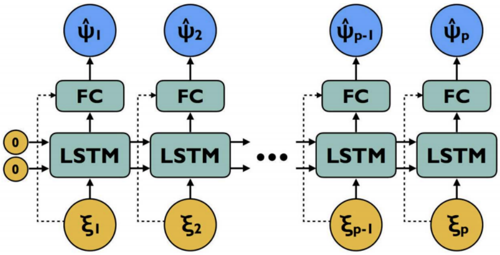


Figure: Architecture of the First Proposed Model

Loss: $L = \sum_{i=1}^{p} w_i ||\hat{\Psi}_i - \Psi_i||^2$, where $w_i = \begin{cases} w_0 & 0 < i \leq p_t \\ 1 & p_t < i \leq p \end{cases}$

LSTM

Problem:

- Input representing the ξ is always accurate regardless of any errors made in predicting the dynamics previously, i.e. the model only learns to predict one step ahead.
- This is undesirable especially for chaotic systems where errors tend to grow exponentially.

(II)

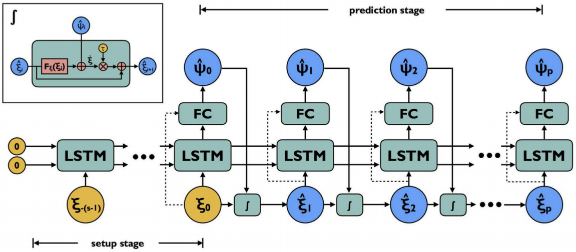


Figure: Architecture of the Second Proposed Model

LSTM

- Loss for the second model:
  $L = \sum_{i=1}^{p} w_i ||\hat{\Psi}_i + F_\xi(\hat{\xi}_i) - \dot{\xi}_i||^2$, where $w_i = \gamma^{i-1}$, $0 < \gamma < 1$.
- Use first model for pre-training, use second model for fine-tuning and prediction.

An alternative to the method proposed is fully data-driven modeling, which simply make predictions of ξ based on previous observed ξs. It is shown that this fully data-driven model is worse in the numerical experiments.

Table of Contents

CDV

$$\dot{x}_1 = \gamma_1^* x_3 - C(x_1 - x_1^*), \qquad \dot{x}_2 = -(\alpha_1 x_1 - \beta_1) x_3 - C x_2 - \delta_1 x_4 x_6,$$

$$\dot{x}_3 = (\alpha_1 x_1 - \beta_1) x_2 - \gamma_1 x_1 - C x_3 + \delta_1 x_4 x_5, \qquad \dot{x}_4 = \gamma_2^* x_6 - C(x_4 - x_4^*) + \varepsilon(x_2 x_6 - x_3 x_5),$$

$$\dot{x}_5 = -(\alpha_2 x_1 - \beta_2) x_6 - C x_5 - \delta_2 x_4 x_3, \qquad \dot{x}_6 = (\alpha_2 x_1 - \beta_2) x_5 - \gamma_2 x_4 - C x_6 + \delta_2 x_4 x_2,$$

where the model coefficients are given by

$$\alpha_m = \frac{8\sqrt{2} m^2 (b^2 + m^2 - 1)}{\pi (4m^2 - 1)(b^2 + m^2)}, \quad \beta_m = \frac{\beta b^2}{b^2 + m^2},$$

$$\delta_m = \frac{64\sqrt{2}}{15\pi} \frac{b^2 - m^2 + 1}{b^2 + m^2}, \quad \gamma_m^* = \gamma \frac{4\sqrt{2} m b}{\pi (4m^2 - 1)},$$

$$\varepsilon = \frac{16\sqrt{2}}{5\pi}, \quad \gamma_m = \gamma \frac{4\sqrt{2} m^3 b}{\pi (4m^2 - 1)(b^2 + m^2)},$$

Figure: CDV Equations

$(x_1^*, x_4^*, C, \beta, \gamma, b) = (0.95, -0.760955, 0.1, 1.25, 0.2, 0.5)$. Use 10000 trajectories, 80% for training, 10% for validation, 10% for test. $n_{LSTM} = 1$, $n_{FC} = 16$. Try to predict $p = 200$ time steps ahead, with $\delta t = 0.01$.

CDV

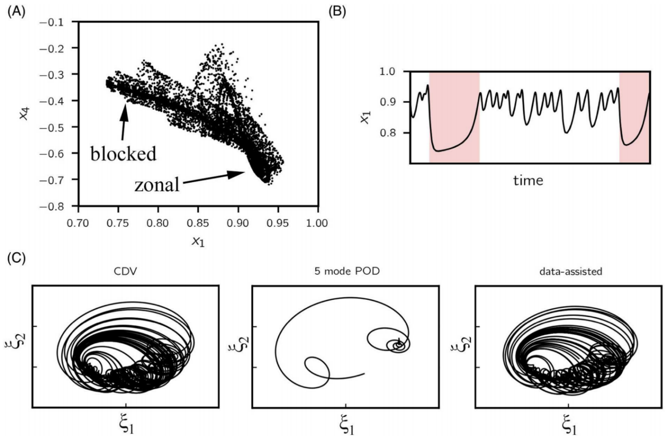


Fig 3. CDV system. (A) 10^4 points sampled from the CDV attractor, projected to (x_1, x_4) plane. (B) Example time series for x_1; *blocked* flow regime is shaded in red. (C) Length-2000 trajectory projected to the first two POD modes (normalized) integrated using the CDV model (left), 5-mode POD projected model (middle) and data-assisted model (right). Despite preserving 99.6% of the total variance, the 5-mode projected model has a single fixed point as opposed to a chaotic attractor. Data-assisted model, however, is able to preserve the geometric features of the original attractor.

Results

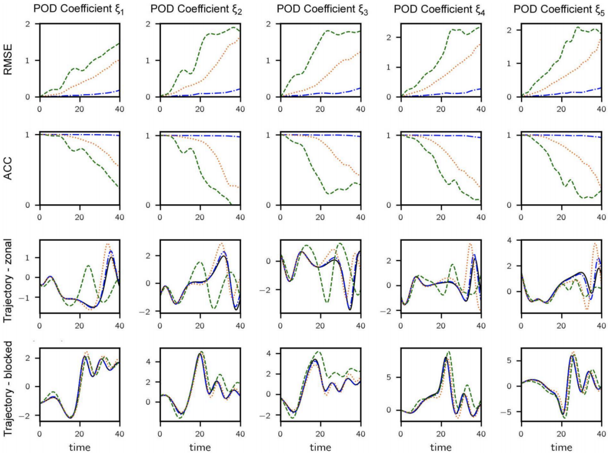


Fig 4. Results for CDV system. (Row 1) RMSE vs. lead time for 5-mode POD projected model (orange dotted), data-assisted model (blue dashdotted) and purely data-driven model (green dashed). (Row 2) ACC vs. lead time. (Row 3) A sample trajectory corresponding to zonal flow—true trajectory is shown (black solid). (Row 4) A sample trajectory involving regime transition (happening around $t = 20$). For rows 1, 3 and 4, plotted values are normalized by the standard deviation of each dimension.

Intermittent bursts of dissipation in Kolmogorov

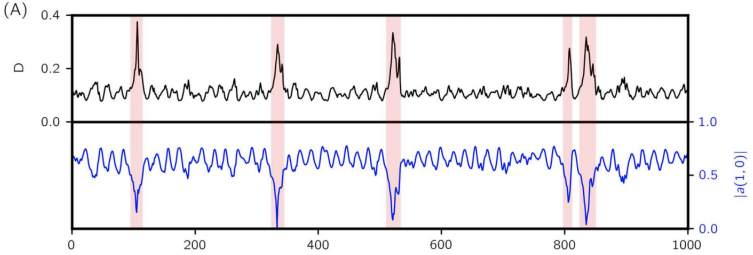
$$\begin{aligned} \partial_t \mathbf{u} &= -\mathbf{u} \cdot \nabla \mathbf{u} - \nabla p + \nu \Delta \mathbf{u} + \mathbf{f} \\ \nabla \cdot \mathbf{u} &= 0 \end{aligned} \tag{17}$$

where $\mathbf{u} = (u_x, u_y)$ is the fluid velocity defined over the domain $(x, y) \in \Omega = [0, 2\pi] \times [0, 2\pi]$ with periodic boundary conditions, $\nu = 1/Re$ is the non-dimensional viscosity equal to reciprocal of the Reynolds number and p denotes the pressure field over Ω. We consider the flow driven by the monochromatic Kolmogorov forcing $\mathbf{f}(\mathbf{x}) = (f_x, f_y)$ with $f_x = \sin(k_f y)$ and $f_y = 0$. $\mathbf{k}_f = (0, k_f)$ is the forcing wavenumber.

$$\begin{aligned} E(\mathbf{u}) &= \frac{1}{|\Omega|} \int_\Omega \frac{1}{2} |\mathbf{u}|^2 \ d\Omega, \\ D(\mathbf{u}) &= \frac{\nu}{|\Omega|} \int_\Omega |\nabla u|^2 \ d\Omega, \\ I(\mathbf{u}) &= \frac{1}{|\Omega|} \int_\Omega \mathbf{u} \cdot \mathbf{f} \ d\Omega \end{aligned} \tag{18}$$

Intermittent bursts of dissipation in Kolmogorov

- The Kolmogorov flow admits a laminar solution $u_x = Re/k_f^2 \sin(k_f y)$, $u_y = 0$. For sufficiently large k_f and Re, this laminar solution is unstable, chaotic and exhibiting intermittent surges in energy input I and dissipation D.
- Study the flow under a particular set of parameters $Re = 40$ and $k_f = 4$ for which there is the occurrence of extreme events.



Due to spatial periodicity, it is natural to examine the velocity field in Fourier space. The divergence-free velocity field **u** admits the following Fourier series expansion:

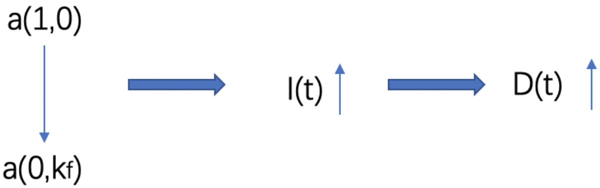
$$\mathbf{u}(\mathbf{x}, t) = \sum_{\mathbf{k}} \frac{a(\mathbf{k}, t)}{|\mathbf{k}|} \begin{pmatrix} k_2 \\ -k_1 \end{pmatrix} e^{i\mathbf{k}\cdot\mathbf{x}} \tag{19}$$

where $\mathbf{k} = (k_1, k_2)$ is the wavenumber and $a(\mathbf{k}, t) = -\overline{a(-\mathbf{k}, t)}$ for **u** to be real-valued. For notation clarity, we will not explicitly write out the dependence on t from here on. Substituting Eq (19) into the governing equations Eq (17) we obtain the evolution equations for a as (more details are presented in S1 Appendix)

$$\dot{a}(\mathbf{k}) = \sum_{\mathbf{p}+\mathbf{q}=\mathbf{k}} i \frac{(p_1 q_2 - p_2 q_1)(k_1 q_1 + k_2 q_2)}{|\mathbf{p}||\mathbf{q}||\mathbf{k}|} a(\mathbf{p}) a(\mathbf{q}) - \nu |\mathbf{k}|^2 a(\mathbf{k}) - \frac{1}{2} i(\delta_{\mathbf{k},\mathbf{k}_f} + \delta_{\mathbf{k},-\mathbf{k}_f}) \tag{20}$$

Intermittent bursts of dissipation in Kolmogorov

- In the reduced model,
  Let $\xi_1 = a(0, k_f)$, $\xi_2 = a(1, 0)$, $\xi_3 = a(1, k_f)$,
  ξ_4, ξ_5, ξ_6 are the conjugate pairs of ξ_1, ξ_2, ξ_3.
- This is because in the interest of predicting I and D, the most revealing interaction to observe is among these nodes.



- However, only 59% energy are contained in these nodes, so a simple linear projection could lead to a bad result.

Intermittent bursts of dissipation in Kolmogorov

- $n_{LSTM} = 70$, $n_{FC} = 38$. Number of time steps in setup stage $= 100$, progressively increase number of time steps in prediction stage from $\{10, 30, 50, 100\}$.
- 100000 trajectories, 80% training, 5% validation, 15% test.
- $n_{epochs} = 1000$, batch size $= 250$, $p_t = 60$, $w_0 = 0.01$, $\gamma = 0.98$.

Intermittent bursts of dissipation in Kolmogorov

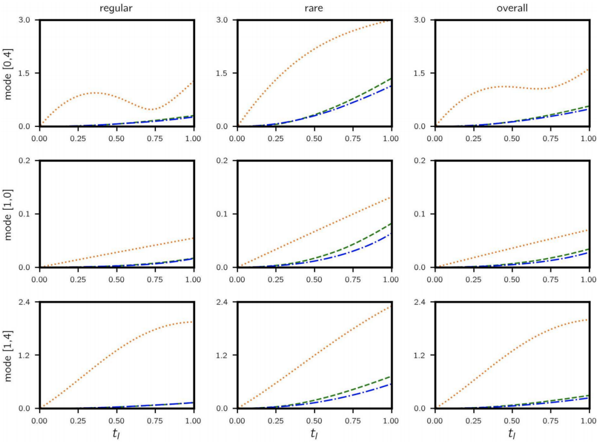


Fig 6. Kolmogorov flow—RMSE vs. time. Errors are computed for 10^4 test trajectories (legend: fully data-driven—green dashed; data-assisted—blue dashdotted; triad—orange dotted). The RMSE in each mode is normalized by the corresponding amplitude $E(\mathbf{k}) = \sqrt{\mathbb{E}[|a(\mathbf{k})|^2]}$. A test trajectory is classified as *regular* if $|a(1,0)| > 0.4$ at $t = 0$ and *rare* otherwise. Performance for regular, rare and all trajectories are shown in three columns. Data-assisted model has very similar errors to those of purely data-driven models for regular trajectories, but the performance is visibly improved for rare events.

Intermittent bursts of dissipation in Kolmogorov

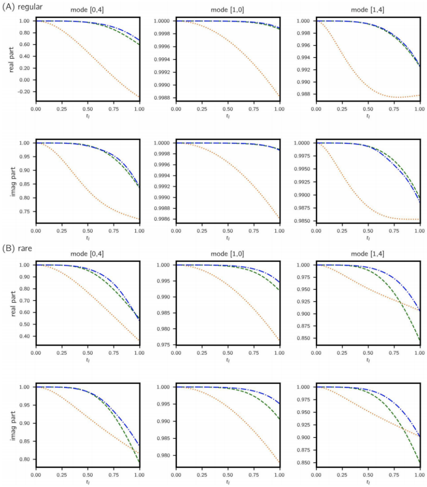


Fig 7. Kolmogorov flow: ACC vs. time. Values are computed for (A) regular and (B) rare trajectories classified from 10^4 test cases. Legend: fully data-driven—green dashed; data-assisted—blue dashdotted; triad dynamics—orange dotted. Real and imaginary parts are treated independently. Similarly to RMSE in Fig 6, improvements in predictions made by the data-assisted model are more prominent for rare events.

Intermittent bursts of dissipation in Kolmogorov

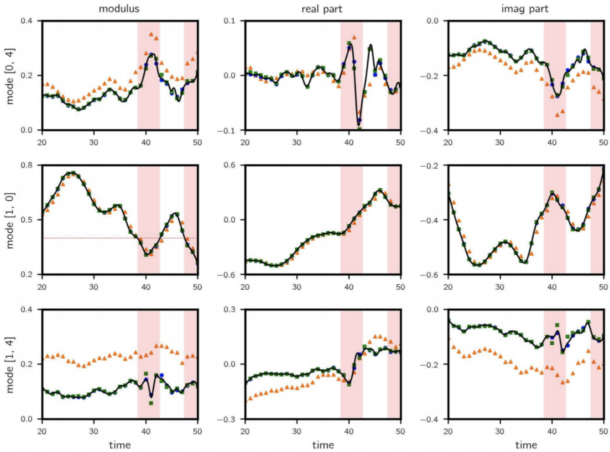


Fig 8. Kolmogorov flow: Predictions along a sample trajectory with lead time = 0.5. Results for the complex modulus (left column), real part (middle column) and imaginary part (right column) of the wavenumber triad are shown. Legend: truth—black solid line; data-assisted—blue circle; triad dynamics—orange triangle; purely data-driven—green square. Rare events are recorded when $|a(1, 0)|$ (left column, mid row) falls below 0.4 (shaded in red). Significant improvements are observed for wavenumbers (0, 4) and (1, 4).

Table of Contents

Conclusion

- Goal of the paper: Dimensionality Reduction of ODEs
- Approach: Data + ODE Prior (Physics-informed LSTM?)
- Testcases on CDV and Navier-Stokes Equations
  Also show the ability to predict rare events
- Possible future research: Singular perturbation?

The End