

Supplementary material

This supplementary material accompanies the research paper: ‘Wang, Y., Kou, J., Noack, B.R. et al. Causal analysis of a turbulent shear flow model’. It contains:

1. Extending GGC analysis to complex, high-dimensional flows exemplified by cavity flows at Reynolds numbers **16000** and **30000**.
2. Significance tests for GGC, including F-test, likelihood ratio test, and permutation test. Their analysis results in the shear flow model and the cavity flow are presented.

1. Extending GGC to complex flows exemplified by cavity flows

1.1 GGC analysis of a cavity flow at $Re=16000$

For complex, high-dimensional flows where the causal structure is not obvious a priori, one approach is to identify a low-dimensional Galerkin model from data (e.g., via POD-Galerkin projection or SINDy regression [1, 2]), and then extract causal relations from the identified low-dimensional governing equations. However, in highly turbulent regimes, such empirical Galerkin models can become unstable or inaccurate, and the complexity of the identified dynamics may limit reliable causal inference.

In contrast, a more broadly applicable approach is to first apply dimensionality reduction (e.g., POD, DMD, or VAE [3]) to the flow field and then perform GGC analysis on the resulting dominant mode coefficients. Theoretically, this approach is justified because the complete set of flow modes constitutes an exact (through high-dimensional) Galerkin system. Applying GGC to the dominant modes thus corresponds to analyzing a truncated version of this system, provided that causal influences from the truncated higher-order modes are negligible.

To demonstrate this, GGC analysis on the 2D post-transient cavity flow data released by Arbabi and Mezić [4] is conducted. The flow data is obtained through a Chebyshev-spectral collocation simulation of an incompressible lid-driven flow in a square domain $[-1, 1]^2$. The Reynolds number is defined as $Re = L_c U_c / \nu = 2 / \nu$, where $L_c = 2$ and $U_c = 1$ are the characteristic length and velocity respectively. Previous studies report that the flow is periodic at $Re=13000$, undergoes a Hopf bifurcation and becomes quasi-periodic at $Re=15000$, and transitions to chaos at $Re>22000$. More details about this simulation can be found in papers of Arbabi and Mezić [4, 5].

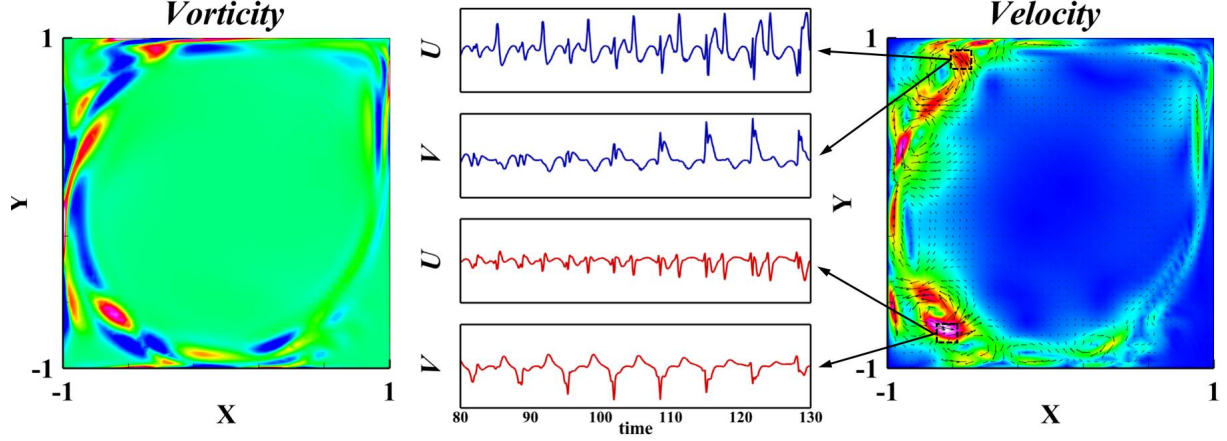


Figure 1. Snapshot of the lid-driven cavity flow at $Re=16000$: vorticity field (left), velocity field with superimposed velocity vectors (right), and time histories of the velocity components at two selected grid points (middle).

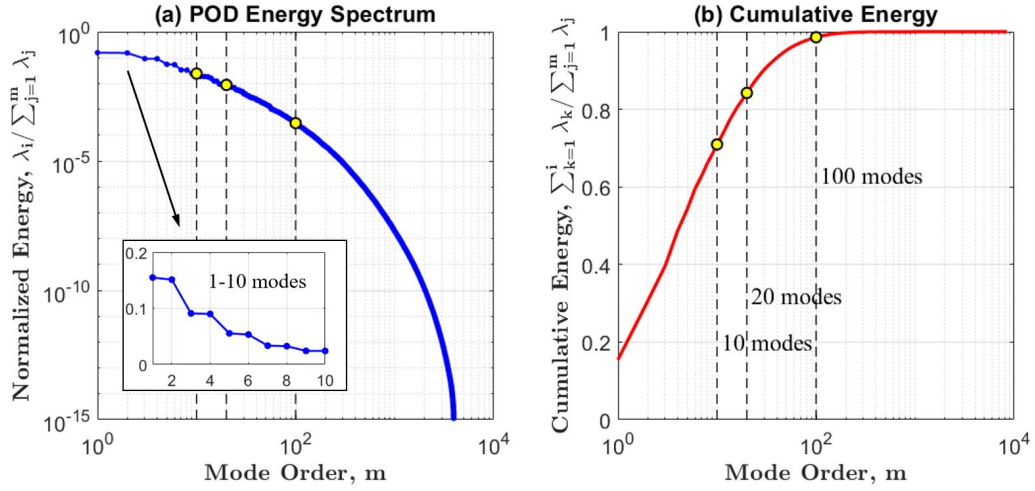


Figure 2. Energy spectrum of the POD modes in the cavity flow of $Re=16000$. (a) Spectrum of the normalized energy of POD modes, defined as $\lambda_i / \sum \lambda_j$. (b) Spectrum of the cumulative normalized energy, defined as $\sum_{k=1}^i \lambda_k / \sum \lambda_j$. The first 10, 20, 100 modes contain 71%, 84%, 99% of the total energy, respectively.

We first apply GGC to the quasi-periodic cavity flow at $Re=16000$, using the 64×64 Chebyshev-grid simulation data from Arbabi and Mezić [5]. POD decomposition is conducted on a dataset comprising 10^4 flow snapshots with a time interval of 0.1. For the two-component velocity field, POD is applied to the vector field by concatenating the u- and v-velocity components into a unified snapshot matrix. The spectrum of POD eigenvalues is shown in Figure 2, where the first 10 modes take up 71% of total energy, capturing the dominant structures of the flow. The resulting POD modes and coefficients are shown in Figure 3. A fast Fourier transformation (FFT) is also conducted on the mode coefficients, which shows that modes 1-4 are dominated by frequency $St=0.155$ and $St=0.252$, modes 5-6 are dominated by a single frequency $St=0.097$, while modes 7-10 exhibit a mixture of frequencies at $St=0.194$, $St=0.349$ and $St=0.408$. The basic frequency vector of the cavity

flow at $Re=16000$ is reported as $St=[0.155, 0.097]$ [4], the dominant frequencies of the first 10 POD modes are linear combinations of those two basic flow frequencies.

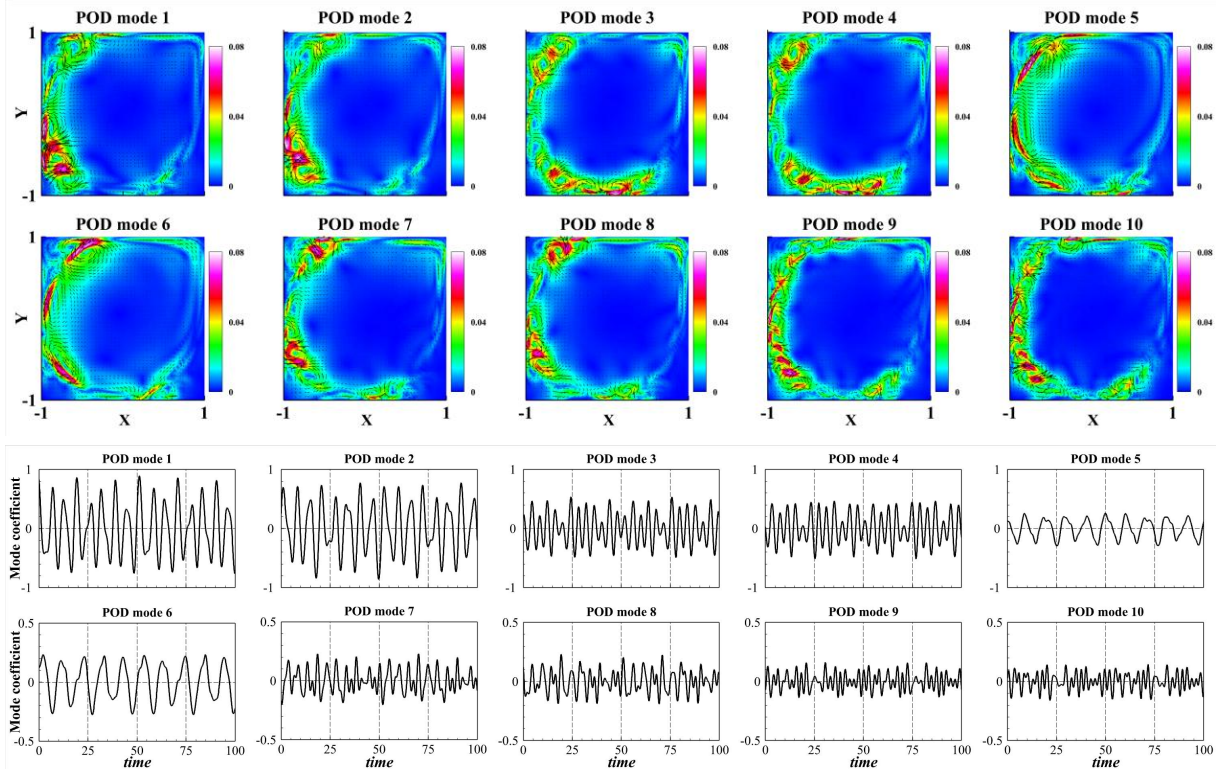


Figure 3. Top: velocity contours of the first 10 POD modes for the lid-driven cavity flow at $Re=16000$, with color indicating velocity magnitude and black arrows depicting the velocity vectors. Bottom: the temporal coefficients of the first 10 POD modes.

GGC is then applied to infer causal relationships between the first 10 POD modes. To determine the appropriate analysis window length, we test a variety of window lengths, the corresponding GGC results and CDI values are shown in Figure 4. Similar to the shear flow model application, the GGC result and CDI value converge as analysis window length increases, and are stabilized beyond 5000 snapshots. This identifies an optimal window length of 5000 snapshots, as evidenced by a clear causal structure shown in Figure 4. Furthermore, this convergence behavior is robust across different numbers of modes considered. A key finding is that the window length required for GGC convergence scales with the number of modes, implying that more data is needed to reliably estimate causality in higher-dimensional systems.

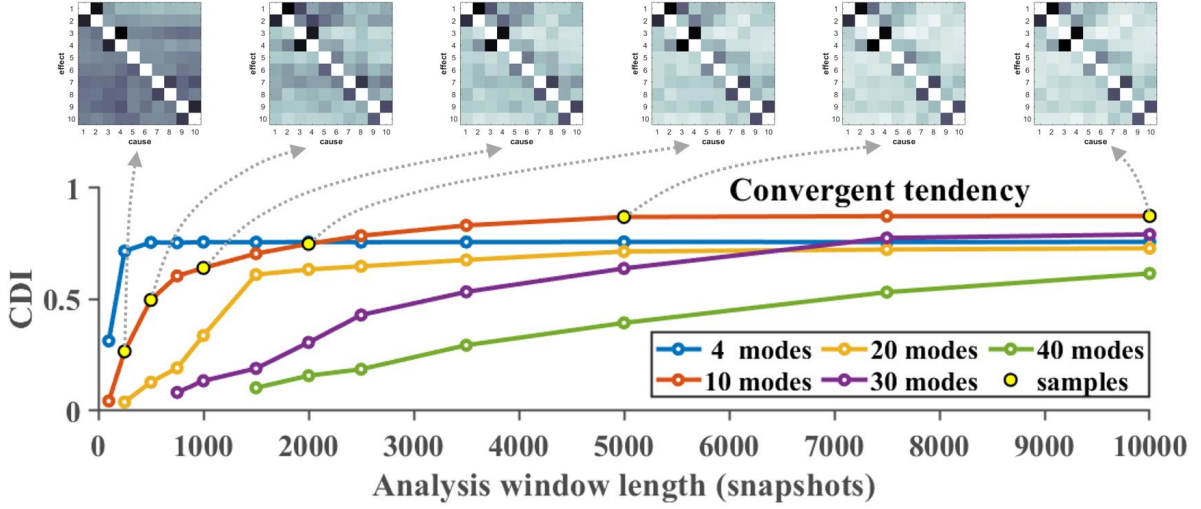


Figure 4. Convergence of the CDI as a function of analysis window length (in snapshots), for GGC analyses using the first 4, 10, 20, 30, and 40 POD modes. In the case of 10 modes, the outcome stabilizes beyond ~ 5000 snapshots ($T = 500$). This convergence becomes slower as the number of modes increases, indicating a need for longer windows in higher-dimensional analyses.

The statistical test reports that significant causalities exist among all modes. Notably, the strongest causal links occur between conjugate mode pairs (e.g., modes $1 \leftrightarrow 2$, $3 \leftrightarrow 4$, ...), as well as from mode 3 to mode 1 and from mode 4 to mode 2. To validate the reliability of the GGC inference, a low-dimension POD-Galerkin model is fitted using the SINDy method proposed by Brunton et.al. [1], whose potential causal relationships are examined and then compared with the GGC results. For model sparsity and interpretability, only the first four POD modes (containing 49% of the total energy) are considered for modeling, and the sparsity threshold is set to $\lambda = 0.1$. The identified Galerkin model is presented below, with its satisfactory representation of the original system shown in Figure 5.

$$\begin{cases} \dot{X}_1 = -1.0837X_2 - 0.3354X_3 + 0.1244X_1X_3 - 0.1329X_1X_4 - 0.2351X_2X_3 \\ \dot{X}_2 = 1.0330X_1 + 0.3303X_4 - 0.1042X_1X_3 - 0.3353X_1X_4 - 0.2256X_2X_3 + 0.1908X_2X_4 \\ \dot{X}_3 = 0.1174X_1 + 1.5265X_4 - 0.1271X_2X_3 - 0.1572X_4^2 \\ \dot{X}_4 = -0.1273X_2 - 1.5064X_3 + 0.1071X_1X_2 - 0.1091X_2X_4 + 0.3153X_3^2 + 0.3224X_4^2 \end{cases}$$

Although all modes seem to influence each other in the equations, we can see strong dependence between conjugate modes (mode $1 \leftrightarrow 2$, $3 \leftrightarrow 4$) through linear interactions, and cross-pair influence (mode $1 \leftrightarrow 3$, $2 \leftrightarrow 4$) from both through linear and nonlinear terms, with comparatively stronger effects from higher-order to lower-order modes. These findings align with the GGC result, supporting the method's validity.

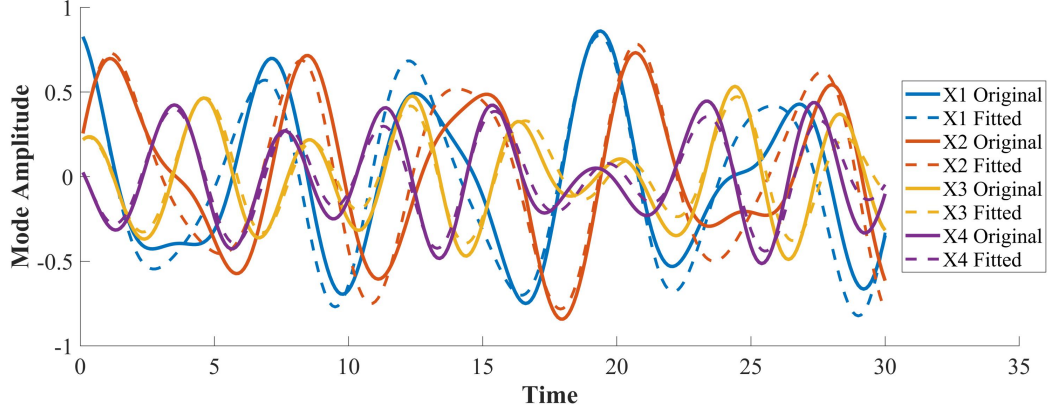


Figure 5. Low-dimensional Galerkin model identified via SINDy regression from the first 4 POD modes of cavity flow at $Re=16000$. The regression threshold is set to 0.1 to enhance sparsity, the resulting model demonstrates reasonable agreement with the original system.

One of the key assumptions underlying Granger causality is fully observed system, which requires no unmeasured confounders [6]. Since GGC is essentially a Granger causality test applied to a low-dimensional Galerkin projection of the full system, a natural concern is whether truncating the model by excluding higher-order POD modes may bias the causal inference. To address this, we performed GGC analyses on models constructed with varying numbers of POD modes (Figure 6). The results reveal that the causal structure among the first 10 modes remains highly consistent across analyses involving 10, 20, and 30 modes. This consistency suggests that the exclusion of higher-order dynamics does not significantly compromise the inferred causal relationships among the dominant modes. This robustness can be attributed to the localized nature of the dynamics. As Figure 6 clearly illustrates, strong causal links exist only between neighboring modes, while longer-range couplings are weak or negligible in the low-dimensional subspace. These findings support the use of GGC on reduced-order models and highlight its robustness in capturing the causal structure of multiscale systems.

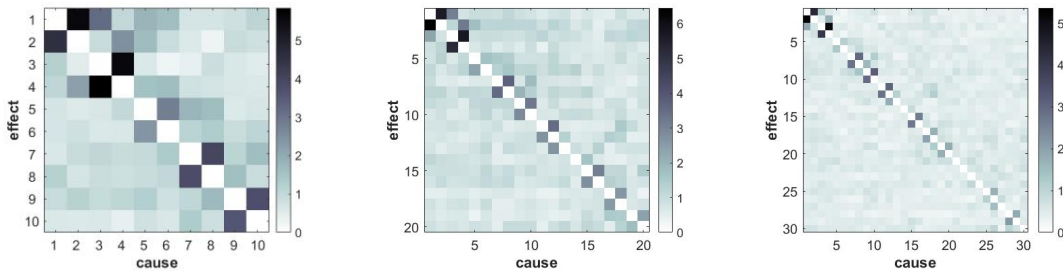


Figure 6. GGC causal maps computed using models constructed from 10, 20, and 30 modes. The consistent causal structure of dominant modes across all cases indicates that the results are robust to the inclusion of higher-order modes, supporting the use of GGC on reduced-order systems.

1.2 GGC analysis of a cavity flow at $Re=30000$

To further validate the GGC method on highly turbulent flows, application on the $Re=30000$ cavity flow is conducted. The simulation is performed on a 128×128 Chebyshev grid. POD decomposition is conducted on 10^4 flow snapshots with a time interval of 0.1. The energy spectrum

shows that the first 10, 20, 100 modes contain 25%, 36%, 66% of the total energy, indicating a highly turbulent flow. The amplitude spectrum of mode coefficients is obtained through FFT, as illustrated in Figure 8. The spectrum demonstrates a multi-frequent characteristic, with the top 3 dominant frequencies being $St=1.080$, $St=1.240$ and $St=0.930$.

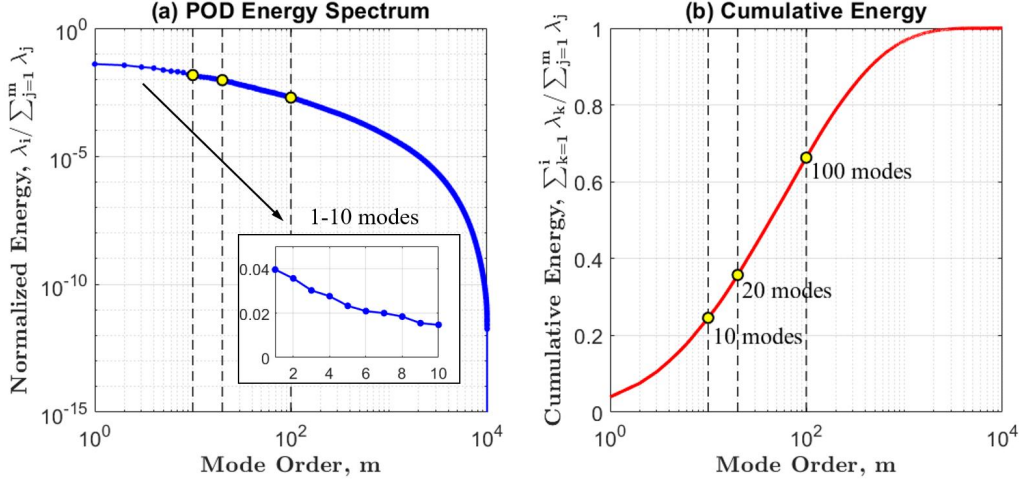


Figure 7. Energy spectrum of the POD modes in the cavity flow of $Re=30000$. (a) Spectrum of the normalized energy of POD modes, defined as $\lambda_i / \sum \lambda_i$. (b) Spectrum of the cumulative normalized energy, defined as $\sum_{k=1}^i \lambda_k / \sum \lambda_j$. The first 10, 20, and 100 modes capture only 25%, 36%, and 66% of the total energy, reflecting the highly turbulent nature of the flow.

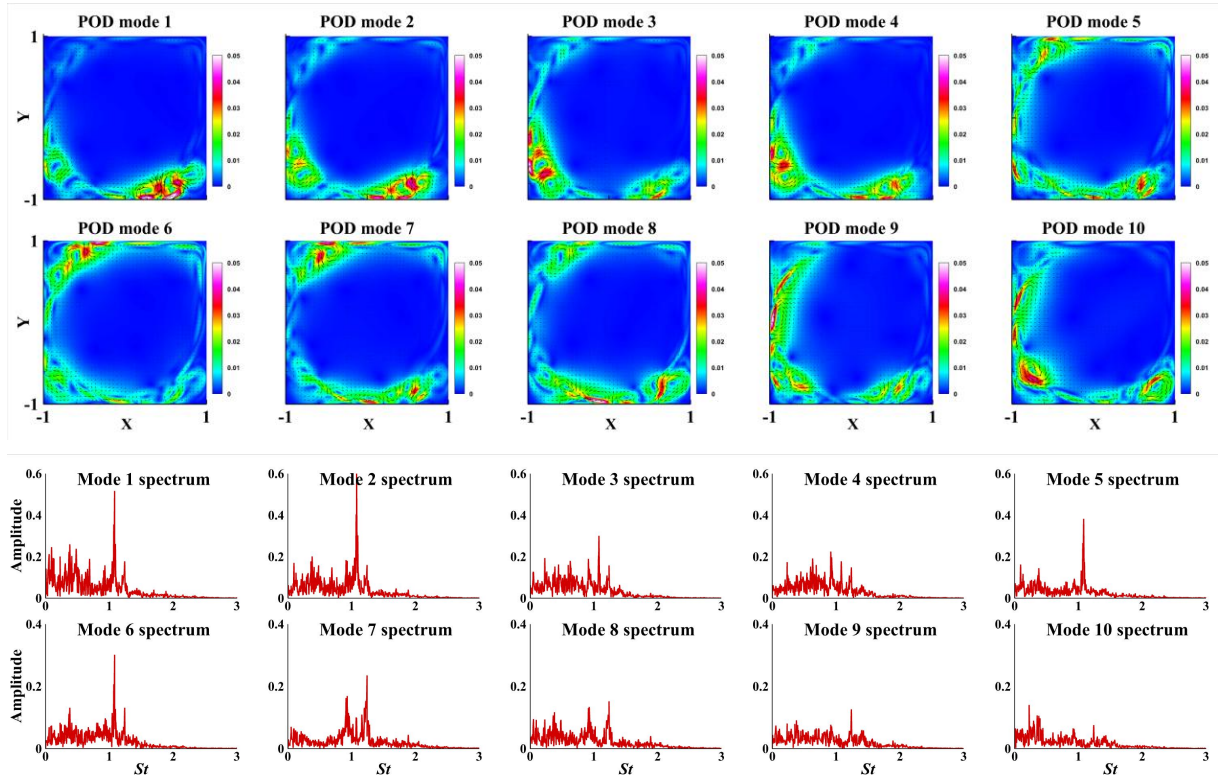


Figure 8. Top: velocity contours of the first 10 POD modes for the lid-driven cavity flow at

$Re=30000$, with color indicating velocity magnitude and black arrows depicting the velocity vectors. Bottom: amplitude spectra of the mode coefficients, obtained via FFT, revealing multi-frequency content with dominant frequencies $St \approx 0.93, 1.08, 1.24$.

As shown in Figure 9, the same convergent tendency of GGC result is observed in cavity flow of $Re=30000$. Compared to the case of $Re=16000$, the convergence process gets slightly slower, indicating more snapshots might be needed to get a reasonable result at higher Reynolds number. Similar to the case of $Re=16000$, consistent GGC results are obtained when including more high-order modes to the reduced-order model, as illustrated in Figure 10. It shows that GGC still demonstrates strong robustness in complex turbulent flow. Due to the multiscale and chaotic nature of turbulence at $Re=30000$, constructing a stable and accurate low-dimensional Galerkin model from data proves challenging, limiting the applicability of equation-based causal analysis in this scenario. The GGC method, however, provides a satisfactory solution for such cases.

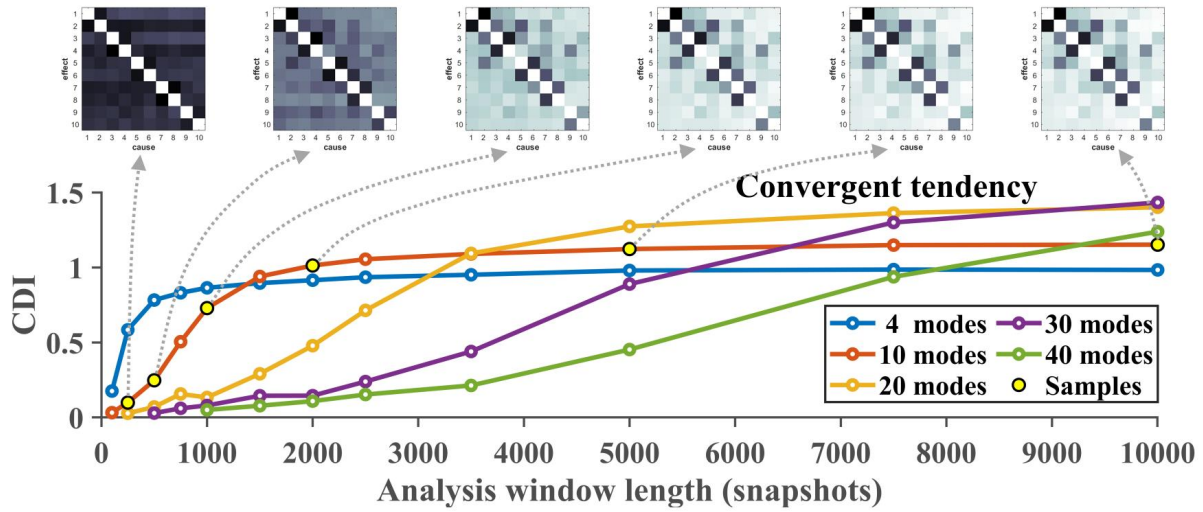


Figure 9. Convergence of the CDI as a function of analysis window length (in snapshots), for GGC analyses using the first 4, 10, 20, 30, and 40 POD modes at $Re=30000$. Convergence is slower than at $Re=16000$, suggesting that more data may be required for reliable inference in highly turbulent regimes.

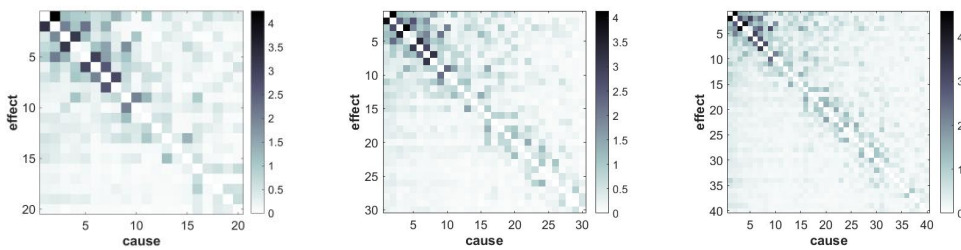


Figure 10. GGC results on models constructed with 20, 30 and 40 POD modes of $Re=30000$ cavity flow. Consistent GGC outcomes of dominant modes are obtained across analyses involving 20, 30 and 40 modes, similar to the case of $Re=16000$.

Moreover, the GGC results for $Re=30000$ are slightly different from the $Re=16000$ case. First, among the first 10 modes, new causal links emerge between modes 2 and 3 and between modes 6

and 7, bridging adjacent conjugate mode pairs. Second, weaker causal influences from higher-order modes toward the dominant modes appear (visible as small gray boxes near the top of the causal map), which is not seen at $Re=16000$. This resembles a reverse cascade in turbulence, where smaller coherent structures influence larger ones. However, a physical interpretation of these observations lies beyond the scope of the present work, which mainly focuses on the validation of the proposed method. Future works are needed to investigate the underlying mechanisms of these causal results.

References:

1. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* 113, 3932–3937 (2016). <https://doi.org/10.1073/pnas.1517384113>
2. Loiseau, J.-C., Brunton, S.L.: Constrained sparse Galerkin regression. *J. Fluid Mech.* 838, 42 – 67 (2018). <https://doi.org/10.1017/jfm.2017.823>
3. Solera-Rico, A., Sanmiguel Vila, C., Gómez-López, M., Wang, Y., Almashjary, A., Dawson, S.T.M., Vinuesa, R.: β -Variational autoencoders and transformers for reduced-order modelling of fluid flows. *Nat. Commun.* 15, 1361 (2024). <https://doi.org/10.1038/s41467-024-45578-4>
4. Arbabi, H., Mezic, I.: Spectral analysis of mixing in 2D high-Reynolds flows.
5. Arbabi, H., Mezić, I.: Study of dynamics in post-transient flows using Koopman mode decomposition. *Phys. Rev. Fluids.* 2, 124402 (2017). <https://doi.org/10.1103/PhysRevFluids.2.124402>
6. Shojaie, A., Fox, E.B.: Granger Causality: A Review and Recent Advances. *Annu. Rev. Stat. Its Appl.* 9, 289–319 (2022). <https://doi.org/10.1146/annurev-statistics-040120-010930>

2. Significance test of GGC analysis

F-test and Likelihood Ratio (LR) test are the typical statistical tests for classical Granger causality. Both can be extended to the proposed GGC framework.

The F-test [1] quantifies the significance of causality by evaluating the deviation between the reduced model and the full model. The F-test statistic is

$$F = \frac{(SSR_{red} - SSR_{full}) / p}{SSR_{full} / (T - k)},$$

where p is the number of restrictions (the number of parameters removed in the reduced model). T is the sample size, while k is the the number of parameters in the full model. Under the null hypothesis H_0 (no Granger causality), this F-statistic follows an F-distribution with degrees of $df_1 = p$ and $df_2 = k$. If $P(F > F_{critical} | H_0) < \alpha$, then H_0 is rejected, indicating a Granger causality exists from the predictor variable to the target variable.

In addition to the classical F-test, Granger causality can also be assessed using the Likelihood Ratio (LR) test [2]. The LR test compares the maximized log-likelihoods:

$$LR = T \ln\left(\frac{RSS_{red}}{RSS_{full}}\right)$$

Under the null hypothesis of no Granger causality, the LR statistic asymptotically follows a chi-squared distribution with degrees of freedom p :

$$LR \xrightarrow{d} \chi^2(p) \text{ as } T \rightarrow \infty.$$

If the computed LR value exceeds the critical value $\chi^2_{1-\alpha}(p)$ at significance level α (commonly 0.05), the null hypothesis is rejected, providing statistical evidence that the predictor variable Granger-causes the target variable. This framework can be easily extended to the proposed GGC as well.

However, when applied to complex systems such as the shear flow model or turbulent cavity flow, these two approaches tend to overestimate significance, concluding that causalities exist between all variables (see Figure 11 and Figure 12). This likely arises because their underlying assumptions (e.g., linearity, stationarity, and Gaussian distributed error) are violated in highly nonlinear, chaotic fluid systems.

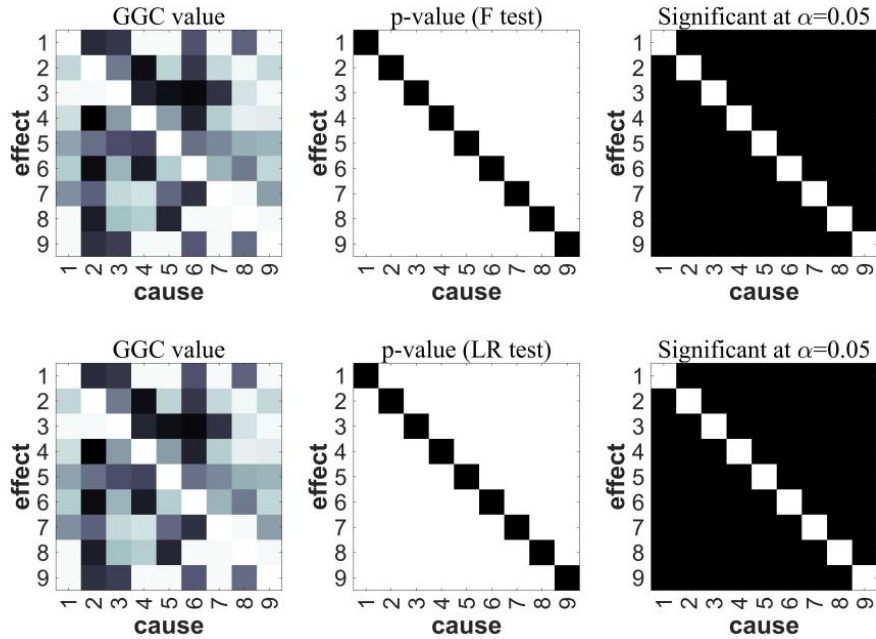


Figure 11. F-test (top) and Likelihood Ratio (LR) test (bottom) results for GGC applied to the turbulent shear flow model. Both tests indicate that all possible causal links exist, over-estimating significance in nonlinear systems.

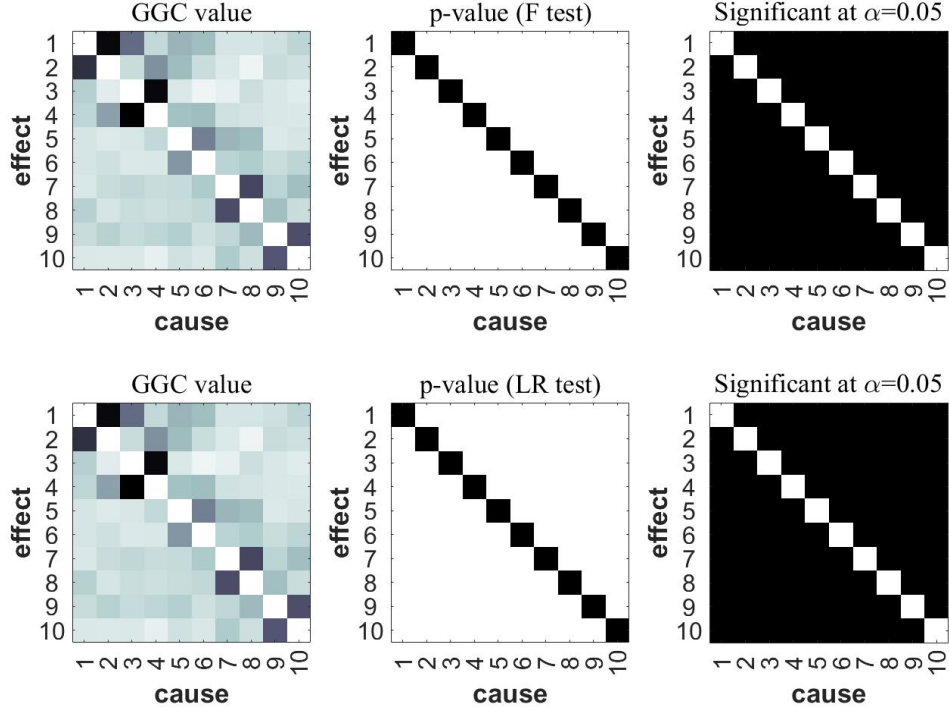


Figure 12. F-test (top) and LR test (bottom) results for GGC applied to the cavity flow at $Re=16000$. Similar to Figure 11, both tests over-estimate significance.

To address this, we employ a permutation test based on time-shifted surrogates [3, 4], which effectively removes secondary and spurious causalities. The core idea is simple: for each candidate cause variable, we generate multiple surrogate time series by applying random cyclic time shifts (i.e., rotating the time series by a random lag). This procedure breaks any potential causality from the cause variable while preserving the intrinsic statistical properties of each series. We then compute GGC for plenty of these shifted surrogate pairs (typically 50-200 permutations) to construct an empirical null distribution under the hypothesis of no Granger causality. The $(1-\alpha)$ -quantile (e.g., the 95th percentile for $\alpha = 0.05$) of this distribution is taken as the significance threshold. If the GGC value from the original data exceeds this threshold, we reject the null hypothesis and conclude that there is statistically significant Granger causality at level α . This approach allows us to reliably identify significant causal links from the quantitative GGC results.

It successfully recovers the true binary causal structure in the shear flow model (see Figure 13). Under 100 permutations, a significance level $\alpha = 0.1$ yields a fully correct causal map,

As shown in Figure 13, using 100 permutations and a significance level of $\alpha = 0.1$, the resulting binary causal map exactly matches the ground-truth causality derived from the governing equations. Satisfactory result can also be obtained with $\alpha = 0.05$, whereas $\alpha = 0.2$ leads to excessive false positives. Thus, the significance level of $0.05 < \alpha < 0.1$ could be a proper range for analysis. Moreover, we observe that results stabilize with about 50 permutations in the shear flow

model, as shown in Figure 14. That suggests robust results can be obtained within reasonable computational cost.

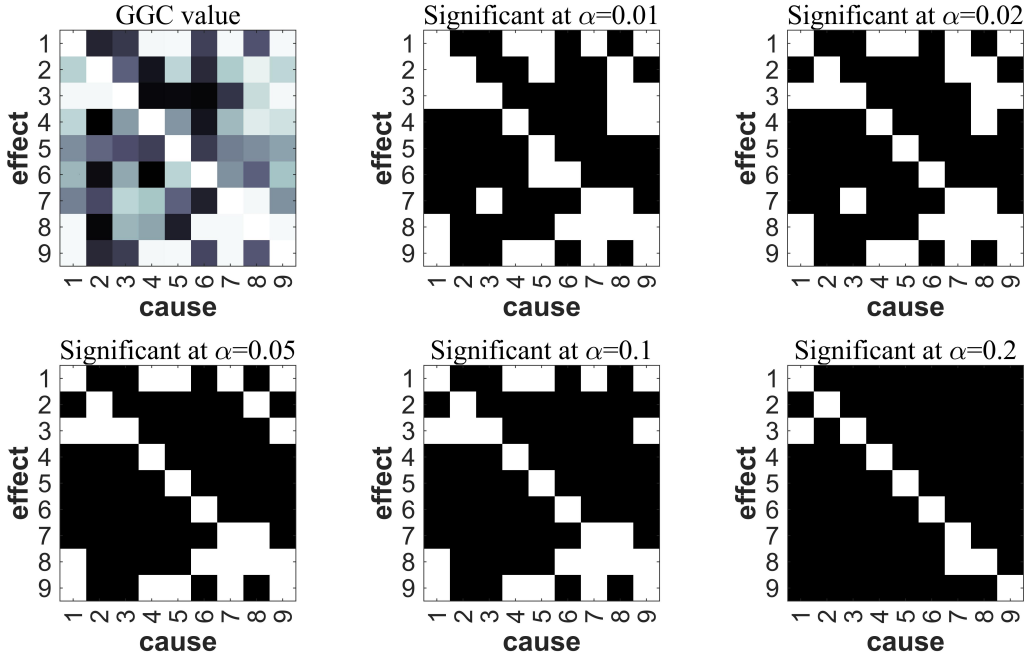


Figure 13. Permutation test results on the turbulent shear flow model (100 surrogates), comparing causal maps at different significance levels α . A complete correct binary causal map (matching ground truth) is recovered at $\alpha = 0.1$.

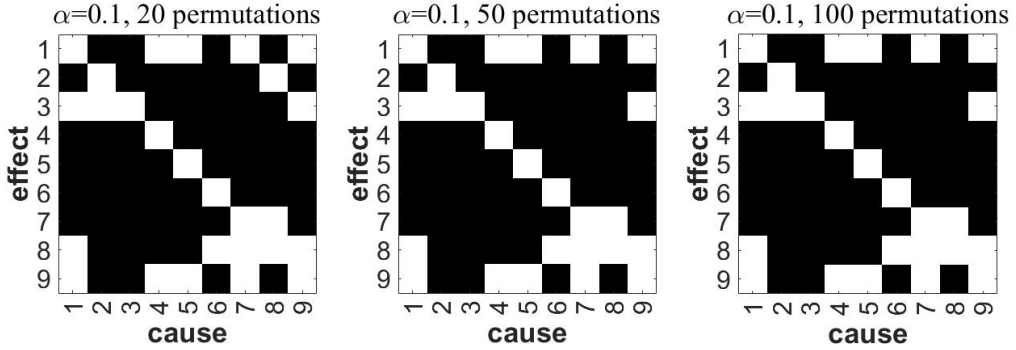


Figure 14. Permutation test results on the turbulent shear flow model at significance level $\alpha = 0.1$, comparing outcomes for 20, 50, and 100 permutations. The causal maps demonstrate method robustness and suggest that about 50-100 permutations is sufficient for stable inference in this system.

When applied to the $Re=16000$ lid-driven cavity flow, the permutation test indicates that nearly all mode pairs exhibit statistically significant causality (Figure 15). This possibly reflects the dense coupling in real turbulent flows, as opposed to the sparsely connected shear flow model. In such cases, we recommend focusing on the magnitude of GGC (rather than significance test) to identify dominant causal effects.

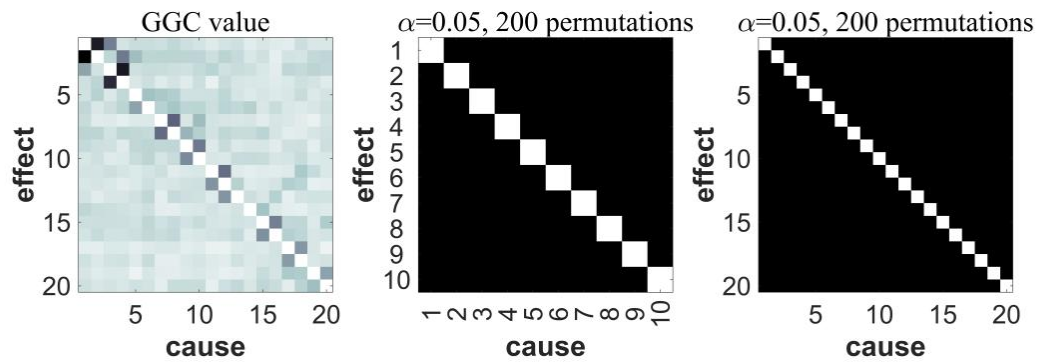


Figure 15. Permutation test results for the cavity flow at $Re=16000$ (200 permutations, $\alpha = 0.05$). All pairwise interactions among the first 10 POD modes are found to be statistically significant, reflecting the dense coupling typical of real turbulent flows.

References:

1. Shojaie, A., Fox, E.B.: Granger Causality: A Review and Recent Advances. *Annu. Rev. Stat. Appl.* 9, 289–319 (2022). <https://doi.org/10.1146/annurev-statistics-040120-010930>
2. Likelihood | SpringerLink, https://link.springer.com/rwe/10.1057/978-1-349-95189-5_874
3. Quian Quiroga, R., Kraskov, A., Kreuz, T., Grassberger, P.: Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E.* 65, 041903 (2002). <https://doi.org/10.1103/PhysRevE.65.041903>
4. Andrzejak, R.G., Kraskov, A., Stögbauer, H., Mormann, F., Kreuz, T.: Bivariate surrogate techniques: Necessity, strengths, and caveats. *Phys. Rev. E.* 68, 066202 (2003). <https://doi.org/10.1103/PhysRevE.68.066202>