PNAS Supporting Information: "Beating the Curse of Dimension with Accurate Statistics for the Fokker-Planck Equation in Complex Turbulent Systems"

Nan Chen and Andrew J. Majda 10.1073/pnas.XXXXXXXXXX

Mutual information

The mutual information (MI) of two random variables X and Y is a measure of the mutual dependence between the two variables (30). MI is more general than the pattern correlation (PC) and MI determines the similarity between the joint distribution $p(X,Y)$ and the products of the factored marginal distribution $p(X)p(Y)$. The definition of the MI is as follows:

$$I(X;Y)=\int_X\int_Y p(X,Y)\log\left(\frac{p(X,Y)}{p(X)p(Y)}\right)dXdY, \quad [1]$$

which is non-negative, i.e.,

$$I(X;Y)\geq 0,$$

and only when X and Y are independent, namely $p(X,Y)=p(X)p(Y)$, the MI is zero.

Different from the PC, the MI contains information about both linear and nonlinear dependence. Therefore, the MI takes into account the information in the non-Gaussian statistics.

In the special case that the two PDFs $p(X)$ and $p(Y)$ are both Gaussian, the MI and PC are linked via the following simple relationship,

$$I(X;Y)=-\frac{1}{2}\ln\left(1-\mathrm{PC}^2(X,Y)\right). \quad [2]$$

In the following, the MI will be adopted to study the dependence between the variables at different spatial grid points in the FitzHugh-Nagumo (FHN) model. Since the PDFs of the variables u_i in the FHN model is highly non-Gaussian, the gap between the MI and PC will be emphasized since it addresses the spatial dependence in the non-Gaussian part.

Constructing the joint PDFs from the efficient statistically accurate algorithms with statistical symmetry

In the main text, we discussed that the statistical symmetry makes use of the samples at different grid points to greatly reduce the number of repeated experiments L. Here we discuss the details of constructing the joint PDFs obtaining from the algorithms with statistical symmetry. Below, we adopt the 1-dimensional case for the convenience of illustration. The method can be easily extended to systems with mutivariables and multidimensions.

Denote $\bar{u}_1,\bar{u}_2,\ldots,\bar{u}_K$ the mean values of the Gaussian ensembles at different grid points and the associated variance are $R_1,\ldots,R_K$. For simplicity, we only take one full run of the system. Therefore, the total number of Gaussian ensembles is K. Clearly, the 1D PDFs $p(u_i)$ at different grid points are the same and are given by

$$p(u_i)=\lim_{k\to\infty}\frac{1}{K}\sum_{k=1}^{K}\mathcal{N}(\bar{u}_k,R_k), i=1,\ldots,K.$$

The limit is taken for statistical intuition while a finite and small K is adopted in practice.

Now we discuss the construction of the joint PDFs. We use the 2D joint PDF $p(u_1,u_2)$ as an illustration. The joint PDF is a Gaussian mixture with K Gaussian ensembles, where the mean of each 2D Gaussian ensemble is

$$\begin{aligned}&\boldsymbol{\mu}_1=\begin{pmatrix}\bar{u}_1\\ \bar{u}_2\end{pmatrix},\quad \boldsymbol{\mu}_1=\begin{pmatrix}\bar{u}_2\\ \bar{u}_3\end{pmatrix},\quad \ldots\\ &\boldsymbol{\mu}_{K-1}=\begin{pmatrix}\bar{u}_{K-1}\\ \bar{u}_K\end{pmatrix},\quad \boldsymbol{\mu}_K=\begin{pmatrix}\bar{u}_K\\ \bar{u}_1\end{pmatrix},\end{aligned} \quad [3]$$

and the covariance matrix is given by

$$\begin{aligned}&\mathbf{R}_1=\begin{pmatrix}R_1 & \\ & R_2\end{pmatrix},\quad \mathbf{R}_2=\begin{pmatrix}R_2 & \\ & R_3\end{pmatrix},\ldots,\\ &\mathbf{R}_{K-1}=\begin{pmatrix}R_{K-1} & \\ & R_K\end{pmatrix},\quad \mathbf{R}_K=\begin{pmatrix}R_K & \\ & R_1\end{pmatrix}.\end{aligned} \quad [4]$$

It is clear from the construction of the ensemble mean in Eq. (3) that the subscript in the second component equals to that of the first component plus one. That is, the first component of each $\boldsymbol{\mu}_k, k=1,\ldots,K$ is treated as u_1 due to the statistical symmetry and the second component is treated as the corresponding u_2 associated with each k. The diagonal covariance matrix in Eq. (4) comes from the fact that each sample point is independent with each other. This is also true and more obvious for the block diagonal conditional covariance. Notably, the diagonal covariance matrix of each ensemble does not mean the correlation between u_1 and u_2 is completely ignored. The correlation is reflected how the points of ensemble means $\boldsymbol{\mu}_k, k=1,\ldots,K$ are distributed .

Scale invariant structures in the stochastic coupled FitzHugh-Nagumo (FHN) model

Here we illustrate the scale invariant structures in the stochastic coupled FitzHugh-Nagumo (FHN) model,

$$\epsilon\frac{du_i}{dt}=u_i-\frac{1}{3}u_i^3-v_i+\sqrt{\epsilon}\delta_1\dot{W}_{u_i}+d_u(u_{i+1}+u_{i-1}-2u_i),\quad i=1,\ldots,N, \quad [5a]$$

$$\frac{dv_i}{dt}=u_i+a+\delta_2\dot{W}_{v_i}. \quad [5b]$$

Quantifying the spatial dependence using mutual information (MI). As shown in the main text, the PDFs of both the variables u_i and v_i in the stochastic coupled FHN model is highly non-Gaussian, where $p(u_i)$ is bimodal and the $p(v_i)$ can be sub-Gaussian. Therefore, the PC, which applies for Gaussian variables, is not a proper measurement to quantify the spatial dependence between the variables at different grid points in the FHN model. To illustrate such insufficiency in the PC and quantify the amount of spatial dependence in a precise way, we introduce the MI and compare it with the PC.

Fig. S1, S2 and S3 show the spatial dependence between u_1 and u_i, where $i = 2, \ldots, N$ in the strongly coherent regime, weakly coherent regime and strongly mixed regime, respectively. Due to the homogenous property, the spatial dependence between $u_{i'}$ and all other u_i with $i' \neq i$ has the same profile as those between u_1 and different u_i. In each figure, the total MI is compared with the MI based on the Gaussian fit PDFs. The latter is equivalent to the PC via the link in Eq. (2), which is also validated in panels (d)-(e) and (i)-(j) in all the three figures.

In the strongly coherent regime (Fig. S1), the total MI (panel (b)) $I(u_1, u_i)$ remains significant $I \geq 0.9$ for all $i = 2, \ldots, 500$. A larger MI is found when i is close to 1, as expected. As comparison, although the MI based on the Gaussian fit PDFs (panel (d)) has a similar profile as the total MI, it decays faster when i increases and its minimum ($\min I_G \approx 0.15$) is much smaller than that of the total MI. This indicates that only around 1/6 information of the spatial dependence between u_1 and u_{250} is from the Gaussian part while the remaining 5/6 information is not reflected in the PC. In fact, with a slight abuse of the concept by plugging $I = 0.9$ into the Eq. (2), the resulting PC is around 0.91 between u_1 and u_{250} while the actual PC is only 0.58. For the other two regimes, the total MI $I(u_1, u_i)$ experiences a rapid decay for $i > 10$ and reaches much smaller values compared with that in the strongly coherent regime. Nevertheless, the gap between the total MI and its Gaussian part remains significant. On the other hand, although the difference between $I(v_1, v_i)$ and $I_G(v_1, v_i)$ is not as large as that associated with u due to the weak non-Gaussianity in v, it is still clear that these is a non-negligible contribution of the spatial dependence from the non-Gaussian part of v.

Scale invariant structures. One of the compelling features of the stochastic coupled FHN model is its scale invariant structure. The scale invariant structure means that the spatial-temporal structures in any given scale has little change as the number of spatial grid points N increases. Fig. S4, Fig. S7 and Fig. S10 show the scale invariant structure in the strongly coherent regime, weakly coherent regime and strongly mixed regime, respectively with N ranging from 50 to 500. A clear indication is shown in columns (b) and (e) with i ranging from 1 to 100 in x-axis. The latter is the zoomed-in results of those in column (d) with $N = 500$. At a fixed grid point $i = 1$, the time series and the associated PDFs are shown in Fig. S5, Fig. S8 and Fig. S11 while the power spectrums and autocorrelation functions are shown in Fig. S6, Fig. S9 and Fig. S12. These figures justify the scale invariant features of the stochastic coupled FHN in all the three regimes.

Finally, in addition to the MI, the following averaged spatial quadratic variations is used to quantify the scale invariant structure,

$$[u]_x = \lim_{T\to\infty} \frac{1}{T} \int_0^T [u(t)]_x dt,$$

where

$$[u(t)]_x = \sum_{s=1}^{N_x-1} \big(u(s+1,t) - u(s,t)\big)^2 \qquad \text{with } t \text{ fixed.}$$

In other words, we first compute the quadratic variation $[u(t)]_x$ that is associated with the 1D curve made up of the spatial points at each fixed time instant t. Then we take the time average of the $[u(t)]_x$. The motivation of introducing this spatial quadratic variation is as follows. The major contribution of the spatial quadratic variation comes from the large intermittent bursts. If the quadratic variations in two simulated fields with different N are comparable, then this is roughly equivalent to saying that the two fields have the same number of large bursts, and it provides the scale invariant structures. Below we also show the averaged number of peaks in spatial direction as a validation of the quadratic variations.

Columns (a) and (c) in Fig. S13 show the averaged spatial quadratic variation as a function of N. Linear functions clearly indicate the scale invariant structures in all the three regimes. As comparison, columns (b) and (d) show the averaged spatial number of peaks, which have the same profile as the averaged spatial quadratic variation. Here the threshold of detecting the peaks are $\tilde{u} = 0$ and $\tilde{v} = 0.5$, respectively. These validate the motivation of adopting the averaged spatial quadratic variation as a more objective measurement for quantifying the scale invariant structures.

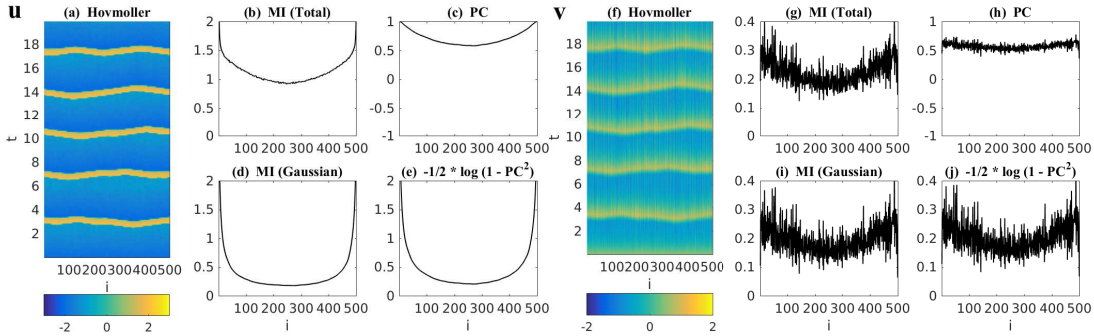

Fig. S1. [FHN model]. Spatial dependence in the **strongly coherent regime**. Panel (a): spatial-temporal structure of u. Panel (b): total mutual information $I(u_1, u_i)$ between u_1 and u_i with $i = 2, \ldots, 500$. Panel (c): pattern correlation PC(u_1, u_i) between u_1 and u_i with $i = 2, \ldots, 500$. Panel (d) mutual information $I_G(u_1, u_i)$ based on the Gaussian fit PDFs of $p(u_1)$ and $p(u_i)$. Panel (e): mutual information with Gaussian assumptions computed from the pattern correlation using the identity in Eq. (2). Panel (f)-(j): similar but for variable v.

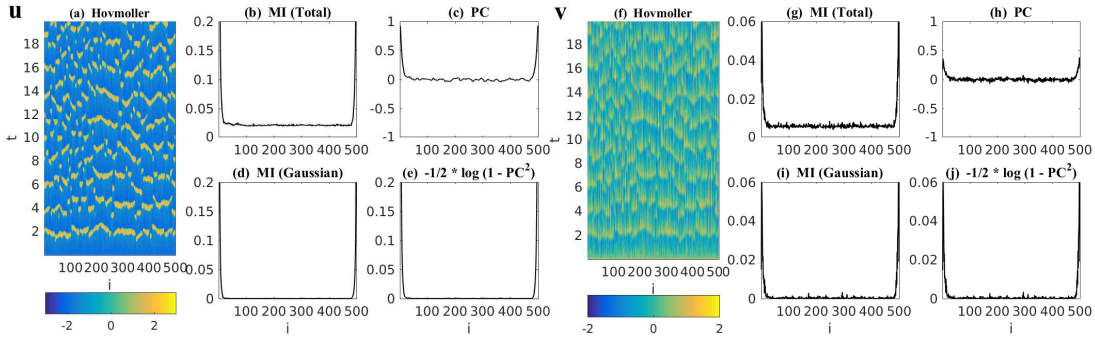

Fig. S2. [FHN model]. Spatial dependence in the **weakly coherent regime**. Panel (a): spatial-temporal structure of u. Panel (b): total mutual information $I(u_1, u_i)$ between u_1 and u_i with $i = 2, \ldots, 500$. Panel (c): pattern correlation PC(u_1, u_i) between u_1 and u_i with $i = 2, \ldots, 500$. Panel (d) mutual information $I_G(u_1, u_i)$ based on the Gaussian fit PDFs of $p(u_1)$ and $p(u_i)$. Panel (e): mutual information with Gaussian assumptions computed from the pattern correlation using the identity in Eq. (2). Panel (f)-(j): similar but for variable v.

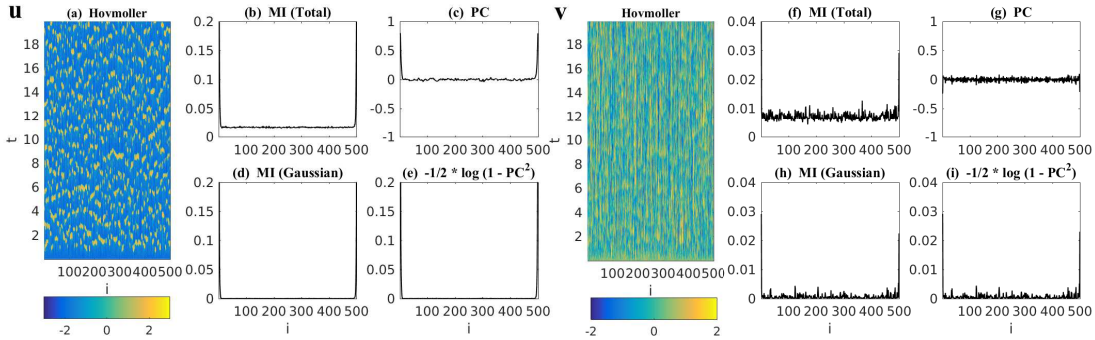

Fig. S3. [FHN model]. Spatial dependence in the **strongly mixed regime**. Panel (a): spatial-temporal structure of u. Panel (b): total mutual information $I(u_1, u_i)$ between u_1 and u_i with $i = 2, \ldots, 500$. Panel (c): pattern correlation PC(u_1, u_i) between u_1 and u_i with $i = 2, \ldots, 500$. Panel (d) mutual information $I_G(u_1, u_i)$ based on the Gaussian fit PDFs of $p(u_1)$ and $p(u_i)$. Panel (e): mutual information with Gaussian assumptions computed from the pattern correlation using the identity in Eq. (2). Panel (f)-(j): similar but for variable v.

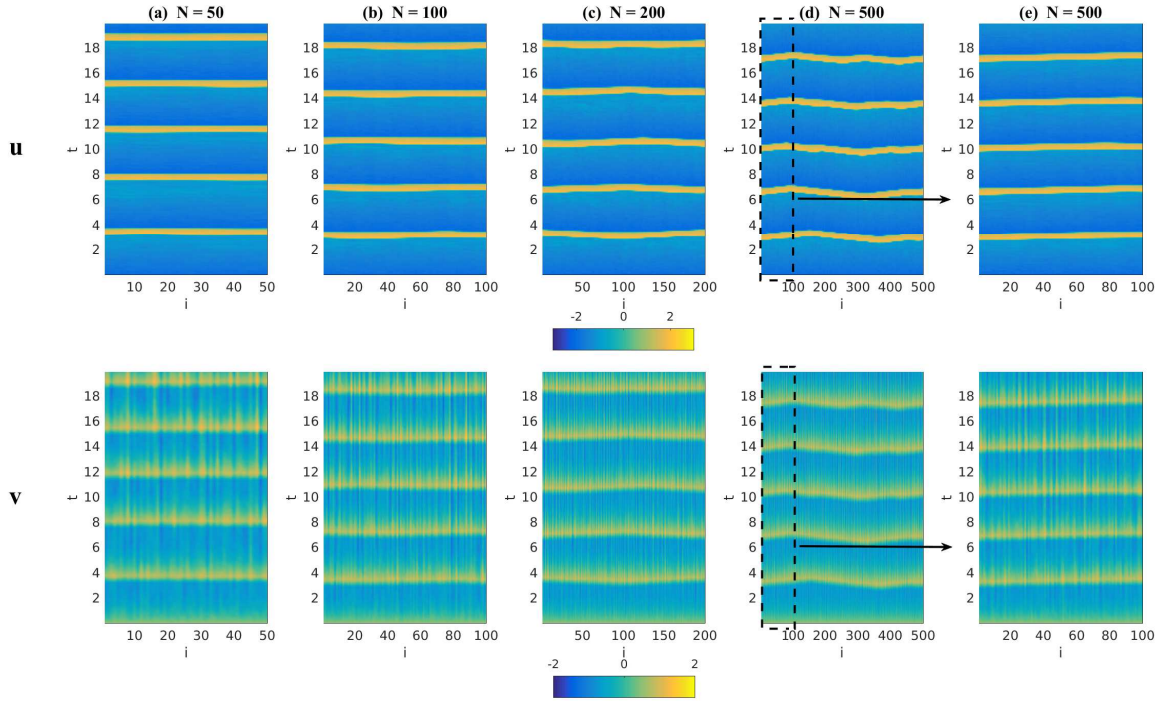

Fig. S4. [FHN model]. Spatial-temporal structures with scale invariant feature in the **strongly coherent regime**, where $\delta_2 = 0.1$ and $d_u = 10$. Columns (a)–(d) show the results with $N = 50, 100, 200$ and 500, respectively. Column (e) is the zoomed-in structures of column (d) with $N = 500$ but it only shows the first 100 spatial grid points, which has similar features as compared with column (b).

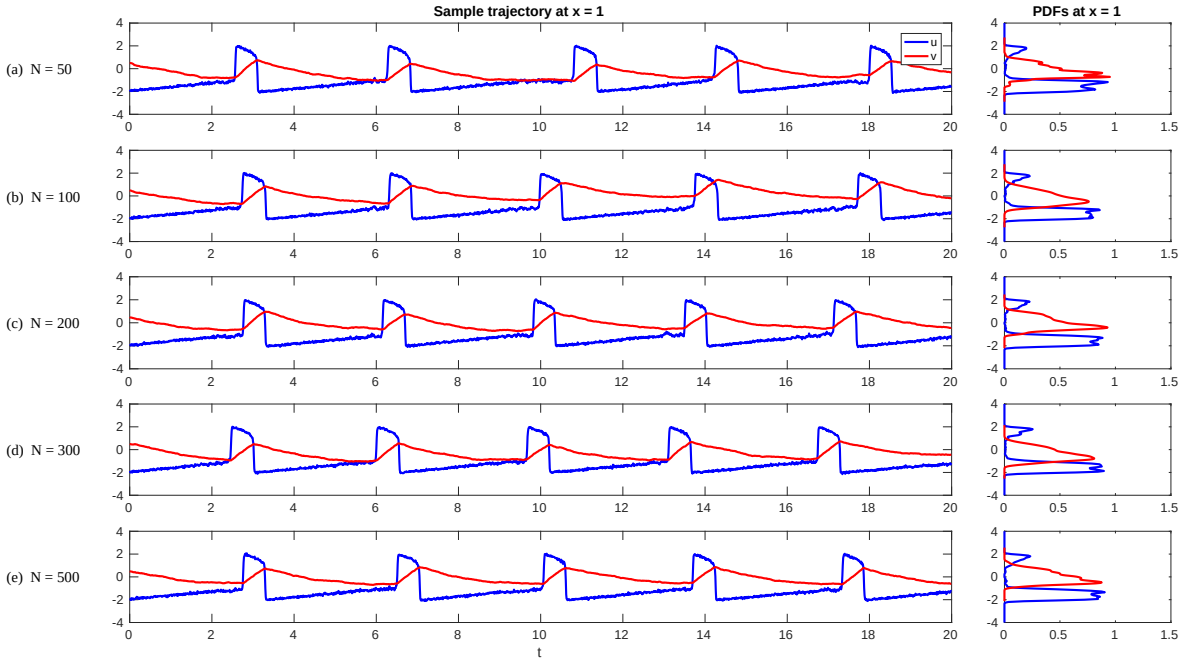

Fig. S5. [FHN model]. Spatial-temporal structures with scale invariant feature in the **strongly coherent regime**, where $\delta_2 = 0.1$ and $d_u = 10$. Rows (a)–(e) show sample time series and the associated PDFs at $x = 1$ for the system with different N. Due to the statistical symmetry, the PDFs at different grid points are the same.

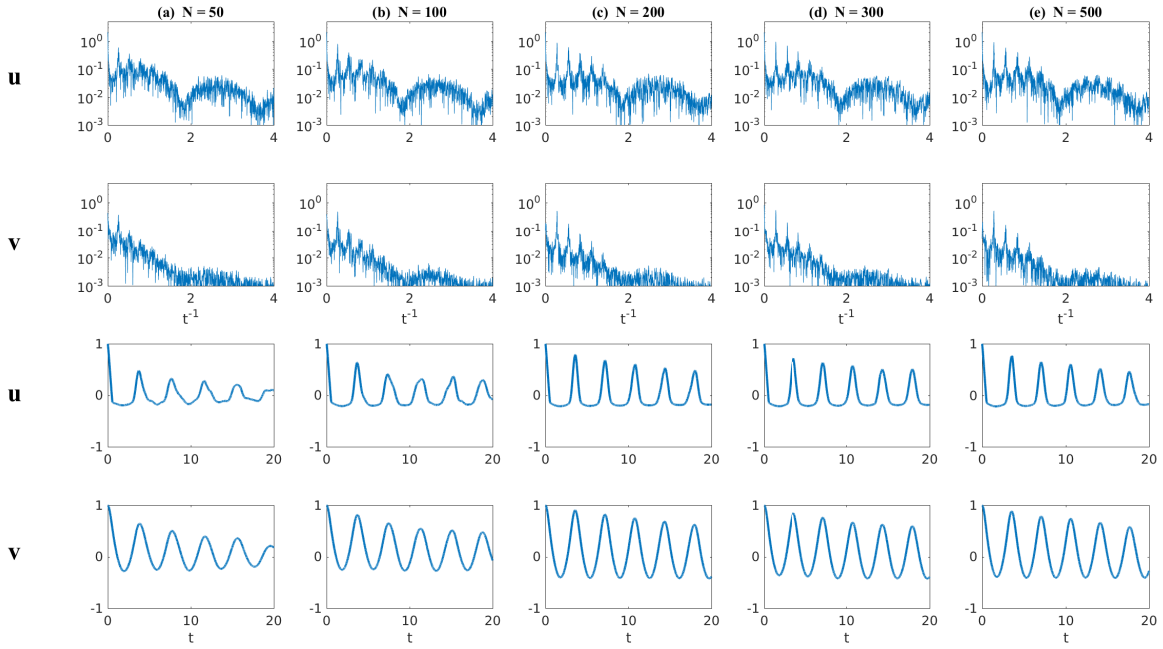

Fig. S6. [FHN model]. Spatial-temporal structures with scale invariant feature in the **strongly coherent regime**, where $\delta_2 = 0.1$ and $d_u = 10$. Top and bottom panels show the power spectrums and autocorrelation functions at $x = 1$ for the system with different N. Due to the statistical symmetry, the PDFs at different grid points are the same.

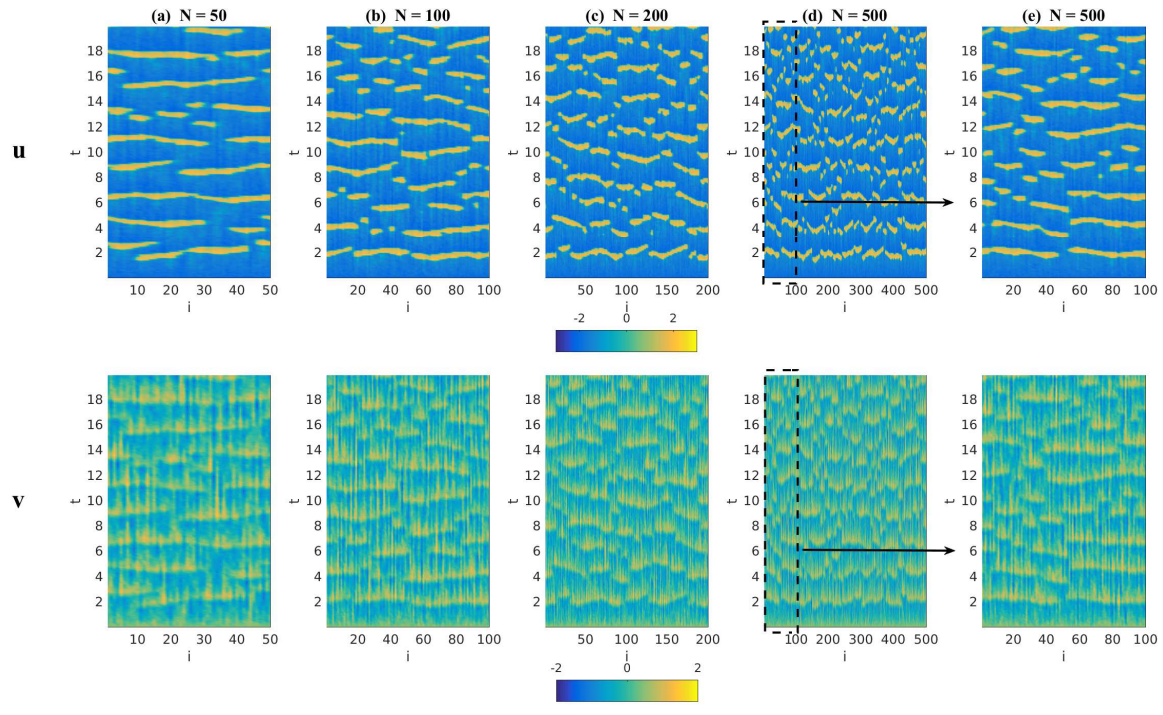

Fig. S7. [FHN model]. Spatial-temporal structures with scale invariant feature in the **weakly coherent regime**, where $\delta_2 = 0.4$ and $d_u = 0.5$. Columns (a)–(d) show the results with $N = 50, 100, 200$ and 500, respectively. Column (e) is the zoomed-in structures of column (d) with $N = 500$ but it only shows the first 100 spatial grid points, which has similar features as compared with column (b).

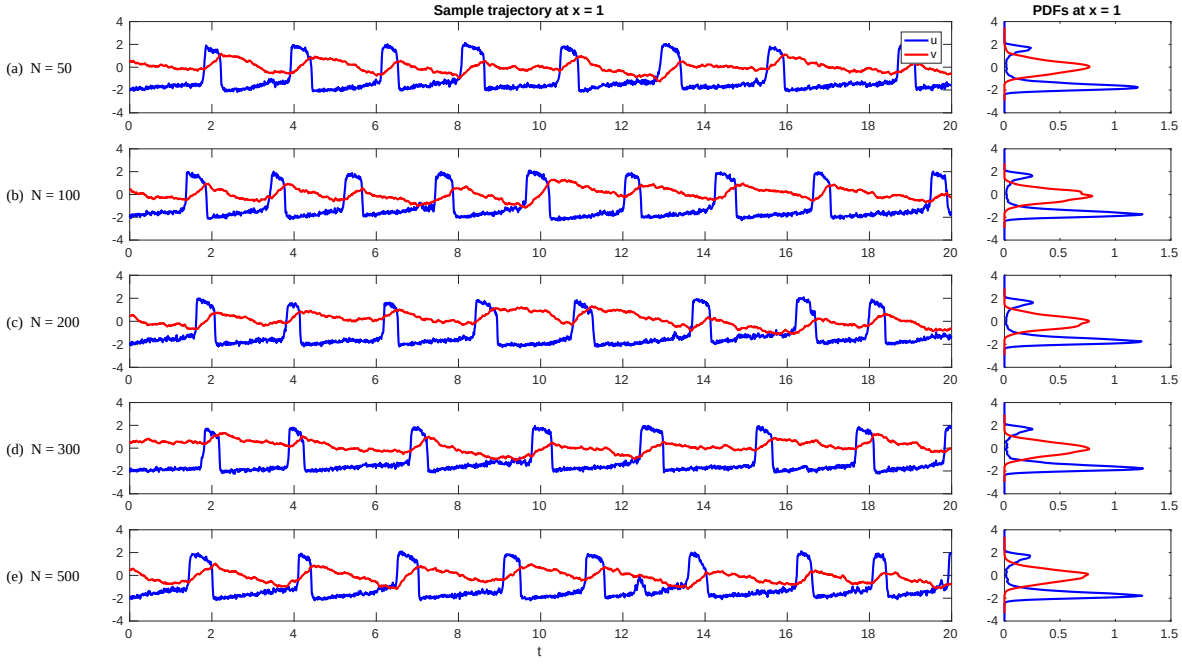

Fig. S8. [FHN model]. Spatial-temporal structures with scale invariant feature in the **weakly coherent regime**, where $\delta_2 = 0.4$ and $d_u = 0.5$. Rows (a)–(e) show sample time series and the associated PDFs at $x = 1$ for the system with different N. Due to the statistical symmetry, the PDFs at different grid points are the same.

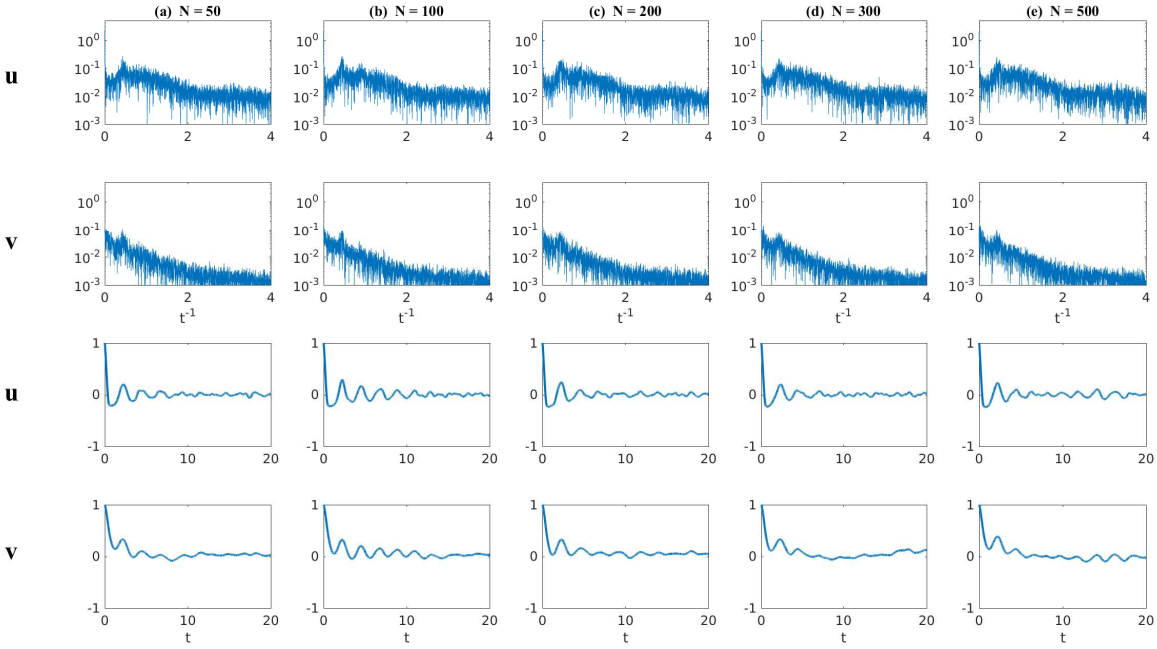

Fig. S9. [FHN model]. Spatial-temporal structures with scale invariant feature in the **weakly coherent regime**, where $\delta_2 = 0.4$ and $d_u = 0.5$. Top and bottom panels show the power spectrums and autocorrelation functions at $x = 1$ for the system with different N. Due to the statistical symmetry, the PDFs at different grid points are the same.

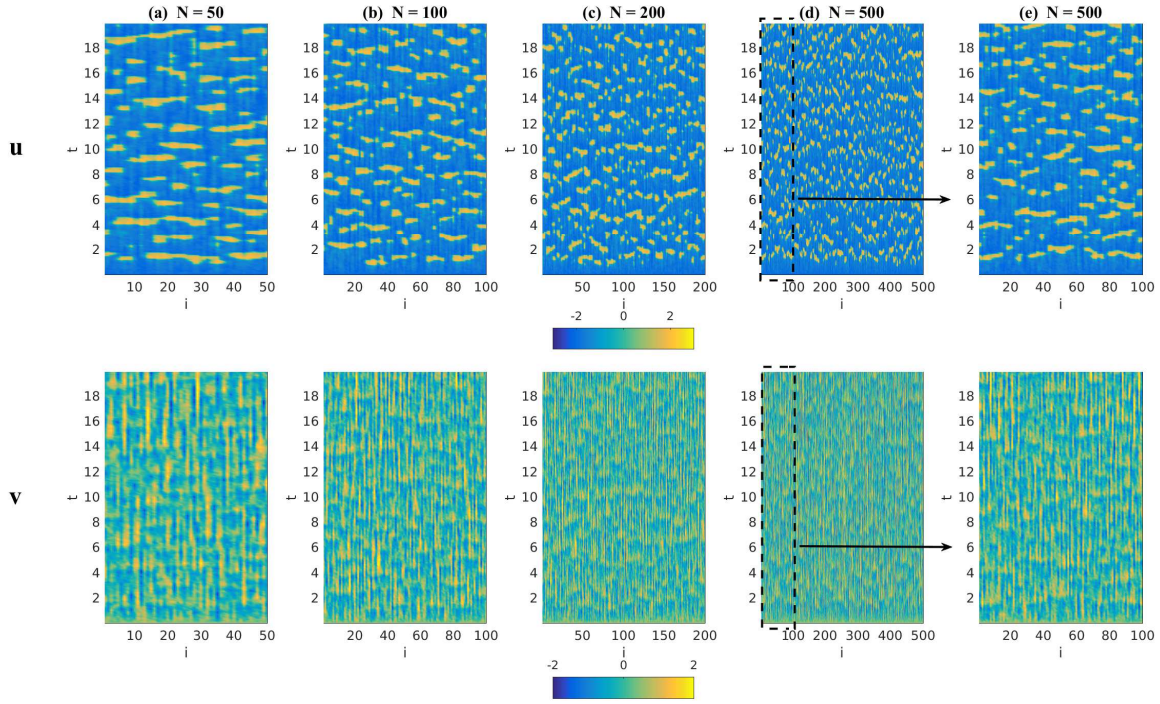

Fig. S10. [FHN model]. Spatial-temporal structures with scale invariant feature in the **strongly mixed regime**, where $\delta_2 = 0.8$ and $d_u = 0.5$. Columns (a)–(d) show the results with $N = 50, 100, 200$ and 500, respectively. Column (e) is the zoomed-in structures of column (d) with $N = 500$ but it only shows the first 100 spatial grid points, which has similar features as compared with column (b).

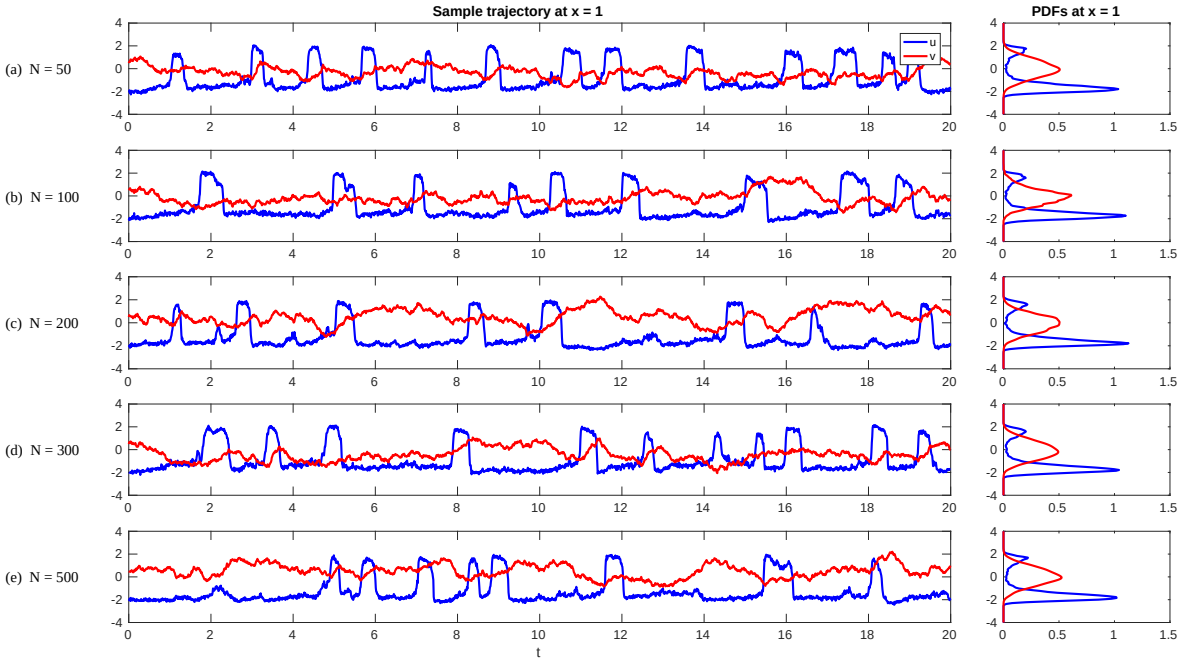

Fig. S11. [FHN model]. Spatial-temporal structures with scale invariant feature in the **strongly mixed regime**, where $\delta_2 = 0.8$ and $d_u = 0.5$. Rows (a)–(e) show sample time series and the associated PDFs at $x = 1$ for the system with different N. Due to the statistical symmetry, the PDFs at different grid points are the same.

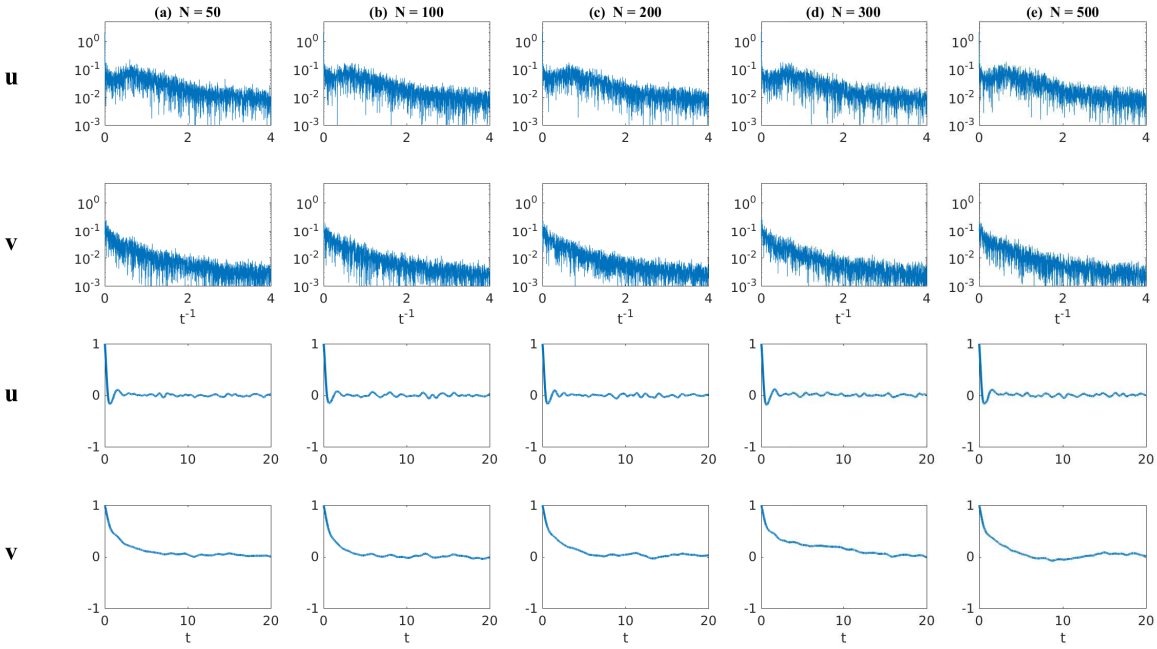

Fig. S12. [FHN model]. Spatial-temporal structures with scale invariant feature in the **strongly mixed regime**, where $\delta_2 = 0.8$ and $d_u = 0.5$. Top and bottom panels show the power spectrums and autocorrelation functions at $x = 1$ for the system with different N. Due to the statistical symmetry, the PDFs at different grid points are the same.

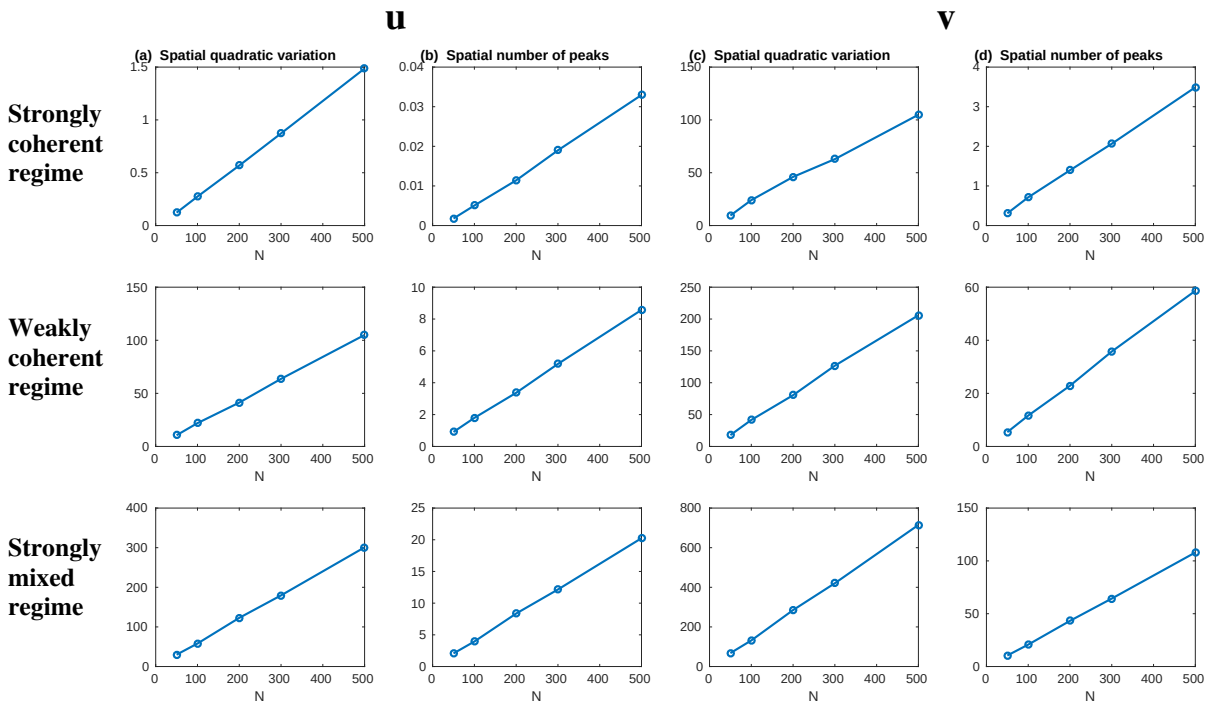

Fig. S13. [FHN model]. Averaged spatial quadratic variation and averaged spatial number of peaks as a function of N in the three different regimes for u (columns (a) and (b)) and v (columns (c) and (d)).

More detailed results of recovering the PDFs of the two-layer inhomogeneous Lorenz 96 model

Recall the two-layer inhomogeneous Lorenz 96 (L96) model:

$$\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) + \sum_{j=1}^{5} \gamma_{i,j} u_i v_{i,j} - \bar{d}_i u_i + F + \sigma_u \dot{W}_{u_i}, \qquad i = 1, \ldots, I, \quad \text{[6a]}$$

$$\frac{dv_{i,j}}{dt} = -d_{v_{i,j}} v_{i,j} - \gamma_j u_i^2 + \sigma_{i,j} \dot{W}_{v_{i,j}}, \qquad j = 1, \ldots, J, \text{ [6b]}$$

where $I = 5$ and $J = 40$ as in the main text and therefore the total number of variables is 240. The parameters are listed in Table S1. With these parameters, the model in Eq. (6) has many desirable properties as in more complicated turbulent systems. For example, both energy and decorrelation decay from large-scale fluctuation variables to small-scale ones while the smaller scales are more intermittent with stronger fat tails in PDFs.

Table S1. Parameters of the two-layer inhomogeneous L96 model in Eq. (6)

σ_u	$\sigma_{k,1}$	$\sigma_{k,2}$	$\sigma_{k,3}$	$\sigma_{k,4}$	$\sigma_{k,5}$
1	0.5	0.2	0.1	0.1	0.1
F	$d_{v_{k,1}}$	$d_{v_{k,2}}$	$d_{v_{k,3}}$	$d_{v_{k,4}}$	$d_{v_{k,5}}$
8	0.2	0.5	1	2	5
$\bar{d}_i$			$\gamma_{i,j} = \gamma_i$		
$1 + 0.7\cos(2\pi i/I)$			$0.1 + 0.025\cos(2\pi i/I)$		

In the coupled system Eq. (6), the nonlinear coupling between $v_{i,j}, j = 1, \ldots, 5$ and u is only through the i-th component of u, namely u_i. This guarantees the block diagonal structure of the conditional covariance matrix, where the total number of entries in the conditional variance matrix is $200^2 = 40,000$ while that of the non-zero entries is $40 \times 5^2 = 1000$. Notably, these 40 blocks of conditional covariance, each being a 5×5 matrix, can be solved in a parallel way. On the other hand, the nonlinear terms $u_{i-1}(u_{i+1} - u_{i-2})$ in Eq. (6a) result in the coupling in the conditional mean of all the $v_{i,j}$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$.

A schematic description of the structure of the two-layer inhomogeneous L96 model is included in Fig. S14. Here $I = 6$ is used for the illustration purpose. All the $I \times J$ variables are coupled with each other, which is reflected in the conditional mean. On the other hand, with each fixed i, the variables $v_{i,j}$ do not interact with $v_{i',j}$ in the conditional covariance matrix with $i' \neq i$.

Fig. S16–S18 show the time series and the associated 1D PDFs of the two-layer inhomogeneous L96 model with $F = 8$ at three different grid points: $i = 1, 11$ and 21, respectively. As shown in the main text, the damping term becomes weaker from $i = 1$ to $i = 21$. Therefore, the variance of u_{21} is larger than that of u_{11} and the variance of u_1 is the smallest. As a result, the non-Gaussian features in $v_{21,j}, j = 1, \ldots, 5$ is the most significant due to the strongest feedback from u_{21} while $v_{1,j}$ with $j = 1, 2, 3$ are nearly Gaussian and $v_{1,4}, v_{1,5}$ do not have as strong fat tails as those in $v_{21,4}, v_{21,5}$. This provides the wave train structures in the spatial-temporal evolution of the variable u in the regions with weaker damping as shown in Fig. 5 of the main text.

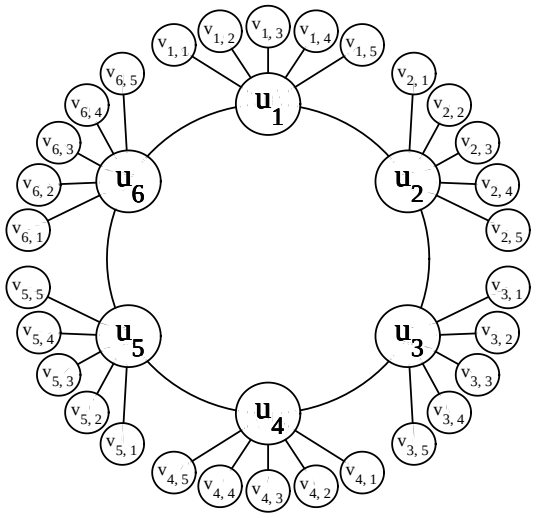

Fig. S14. A schematic description of the structure of the two-layer inhomogeneous L96 model. Here for the illustration purpose, $I = 6$.

Table S2. Pattern correlation between the curves associated truth and recovered 1D marginal statistics of the two-layer inhomogeneous L-96 model as illustrated in Fig. 6 of the main text

	Mean	Variance	Skewness	Kurtosis
u_i	0.97	0.98	0.53	0.78
$v_{i,1}$	0.99	0.99	0.80	0.35
$v_{i,2}$	0.99	0.99	0.87	0.69
$v_{i,3}$	0.99	0.99	0.90	0.80
$v_{i,4}$	0.98	0.99	0.96	0.91
$v_{i,5}$	0.93	0.97	0.98	0.97

Table S2 shows pattern correlations (PCs) between the curves associated with the truth and recovered 1D marginal moments of the two-layer inhomogeneous L-96 model, as illustrated in Fig. 6 of the main text. Significant PCs with values greater than 0.93 are found in the recovered mean and variance for all the variables, namely u_i and $v_{i,j}$ with $j = 1, \ldots, 5$. Notably, even for the higher order moments, namely skewness and kurtosis, the recovered curves are highly correlated with the truth for $v_{i,j}$ with $j = 2, 3, 4, 5$. The variable $v_{i,1}$ is nearly Gaussian and the value of PC in the skewness and kurtosis is not relevant. In addition, the distribution of u_i for all i is nearly symmetric and thus the PC in skewness is irrelevant either.

Fig. S15 compares the recovered 1D PDFs of u_i and $v_{i,j}$ with $j = 1, \ldots, 5$ with the truth at three different grid points: $i = 1, 11$ and 21. With the increase of j, the non-Gaussian features become more significant at all the grid points. The non-Gaussianity also becomes stronger when the damping is weaker. Therefore, the PDFs of the small-scale fluctuation variables at $i = 21$ has the largest skewness and kurtosis. The large-scale variable u is also non-Gaussian and particularly it is bimodal at $i = 11$. The efficient statistically accurate algorithms are able to recover all these PDFs with only $L = 500$ samples.

Fig. S19 and Fig. S20 compare some recovered 2D joint PDFs with the truth at a fixed grid point i, where the results at three different grids $i = 1, 11$ and 21 are shown. In Fig. S19, the joint PDFs between u_i and $v_{i,3}$ and those between u_i and $v_{i,5}$ are shown. In Fig. S20, the joint PDFs of $p(v_{i,1}, v_{i,2})$ and $p(v_{i,5}, v_{i,1})$ are illustrated. The difference in the true PDFs with distinct i is due to the inhomogeneous damping and coupling between u and v variables. Notably, these highly non-Gaussian PDFs are recovered quite accurately by the efficient statistically accurate algorithms with only $L = 500$ samples.

Fig. S21 shows the correlation functions between u_i and u_j (panel (a)), between $v_{i,3}$ and $v_{j,3}$ (panel (b)) and between $v_{i,5}$ and $v_{j,5}$ (panel (c)), where for each fixed i the strongest negative correlation corresponds to $j = i \pm 2$ and the most significant positive correlation happens when $j = i \pm 5$. In Fig. S22 and Fig. S23 we compare the truth and the recovered 2D PDFs for $p(u_i, u_{i+2})$, $p(u_i, u_{i+5})$, $p(v_{i,3}, v_{i+2,3})$ and $p(v_{i,5}, v_{i+2,5})$ with $i = 1, 11$ and 21, respectively. Both the non-Gaussian and correlated features are captured almost perfectly by the efficient statistically accurate algorithms. The inhomogeneity at different grid points are clearly indicated by the recovered PDFs as well.

Finally, we study the error in the recovered PDFs using a quantitative measurement. The natural way to quantify the error in the recovered PDF related to the truth is through an information measure, namely the relative entropy (or Kullback-Leibler divergence) (30). The relative entropy is defined as

$$\mathcal{P}(p(\mathbf{u}), p^M(\mathbf{u})) = \int p(\mathbf{u}) \ln \frac{p(\mathbf{u})}{p^M(\mathbf{u})} d\mathbf{u}, \qquad [7]$$

where $p(\mathbf{u})$ is the true PDF and $p^M(\mathbf{u})$ is the recovered one from the efficient statistically accurate algorithms. This asymmetric functional on probability densities $\mathcal{P}(p, p^M) \geq 0$ measures lack of information in p^M compared with p and has many attractive features. First, $\mathcal{P}(p, p^M) \geq 0$ with equality if and only if $p = p^M$. Secondly, $\mathcal{P}(p, p^M)$ is invariant under general nonlinear changes of variables. Notably, the relative entropy is a good indicator of quantifying the difference in the tails of the two PDFs, which is particularly crucial in the turbulent dynamical systems with intermittency and extreme events. On the other hand, the traditional ways of quantifying the errors, such as the relative error $\|p - p^M\|/\|p\|$, usually underestimate the lack of information in the PDF tails.

The lack of information in a various 1D and 2D PDFs at three different grid points $i = 1, i = 11$ and $i = 21$ are shown in Fig. S24–S26, respectively. The lack of information decays as a function of the number of samples L. Typically, when L is above 200, the curve of the lack of information becomes flat. It is also noticeable that the lack of information has a uniform behavior at different grid points, indicating its robustness.

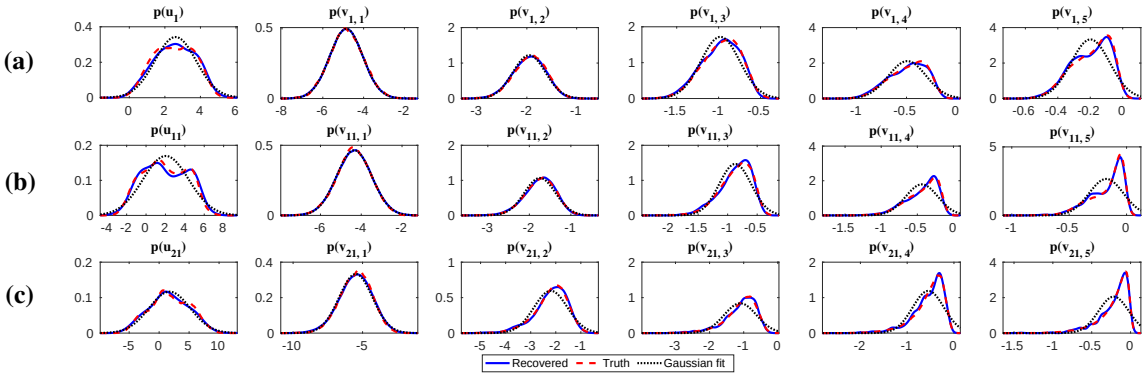

Fig. S15. [Two-layer inhomogeneous L-96 model]. Comparison of the 1D marginal PDFs of u_i and $v_{i,j}$ with $j = 1, \ldots, 5$ in the two-layer L-96 model Eq. (6). Rows (a), (b) and (c) show the PDFs at grids $i = 1, 11$ and 21, respectively.

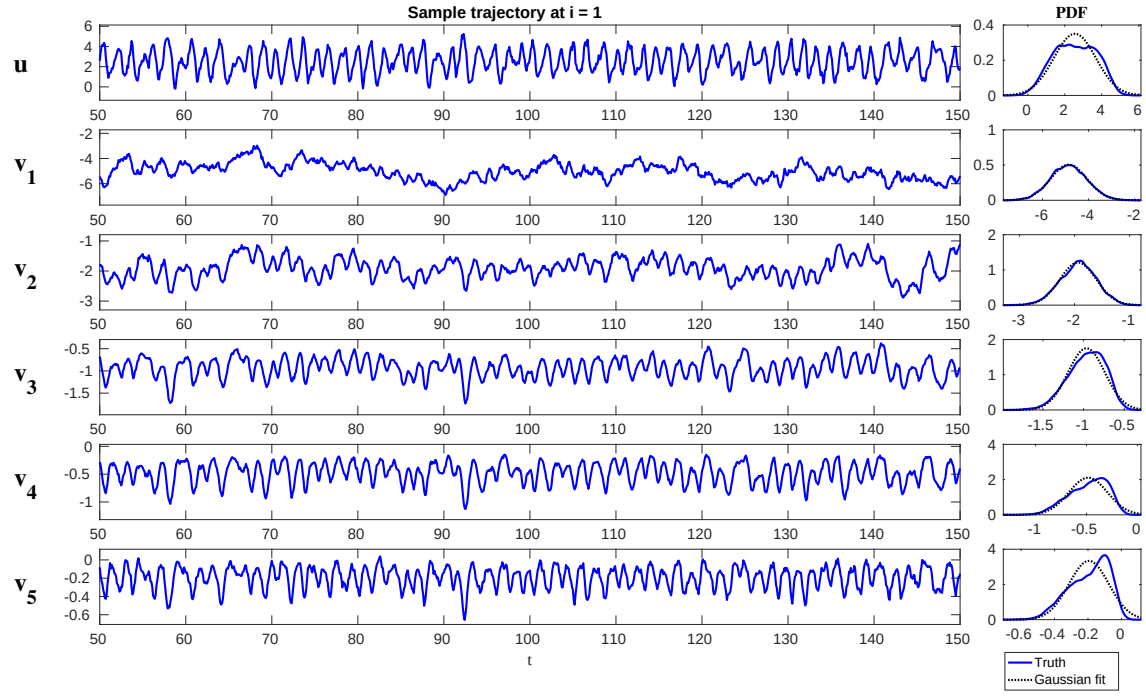

Fig. S16. [Two-layer inhomogeneous L-96 model]. Sample trajectories and PDFs at $i = 1$.

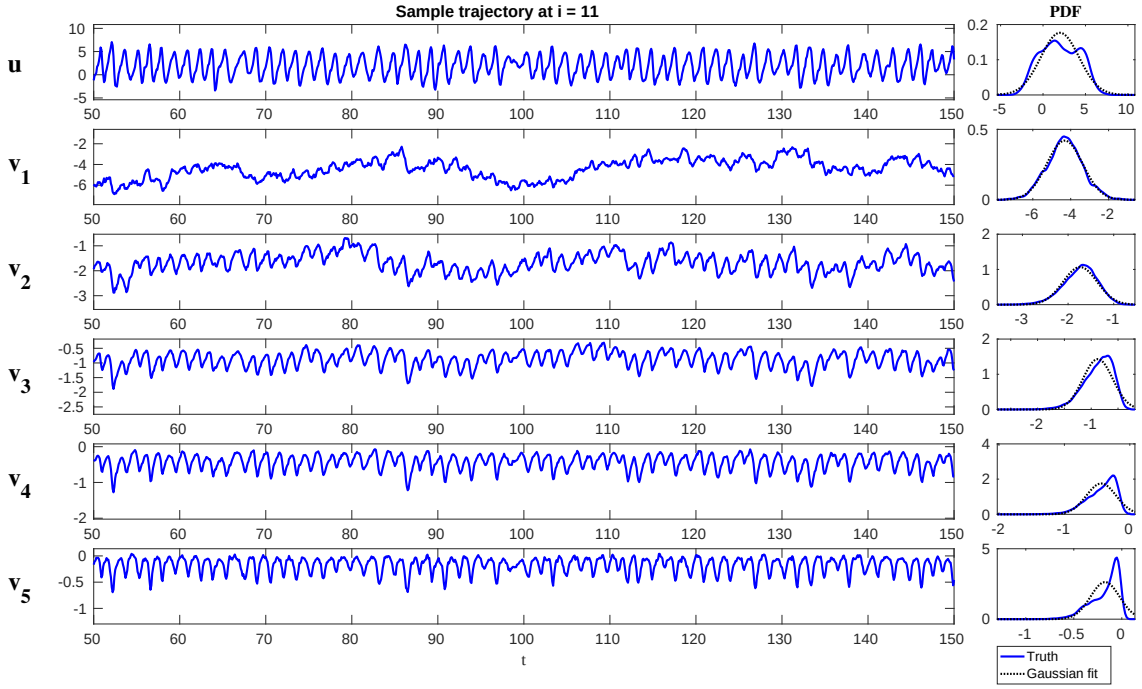

Fig. S17. [Two-layer inhomogeneous L-96 model]. Sample trajectories and PDFs at $i = 11$.

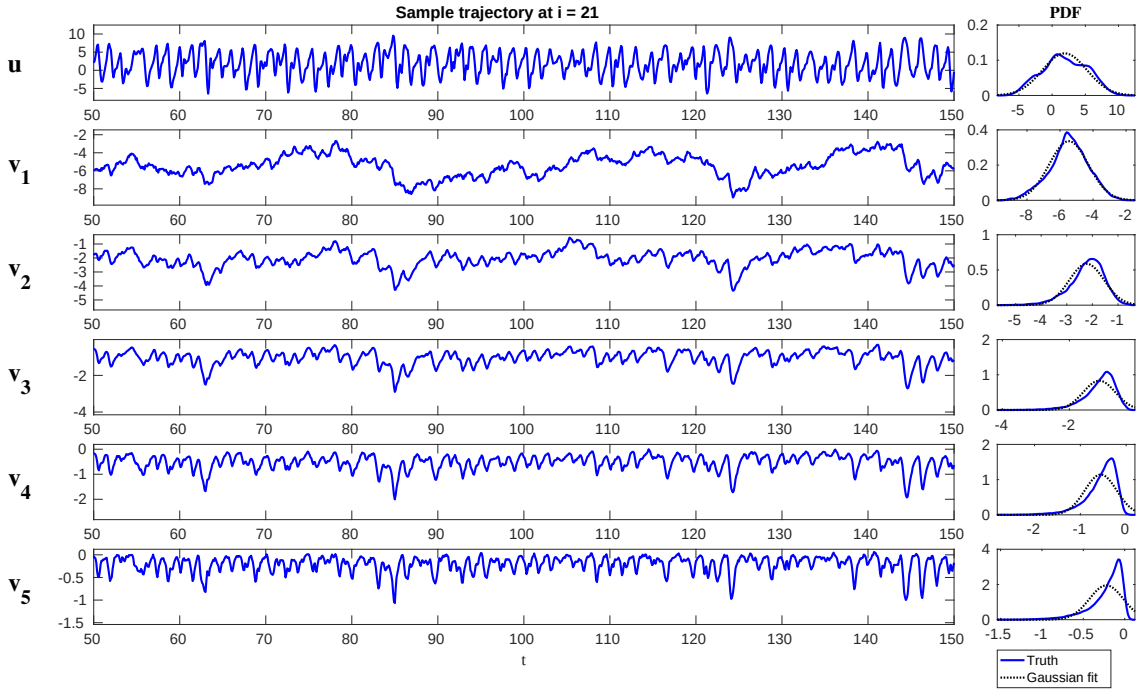

Fig. S18. [Two-layer inhomogeneous L-96 model]. Sample trajectories and PDFs at $i = 21$.

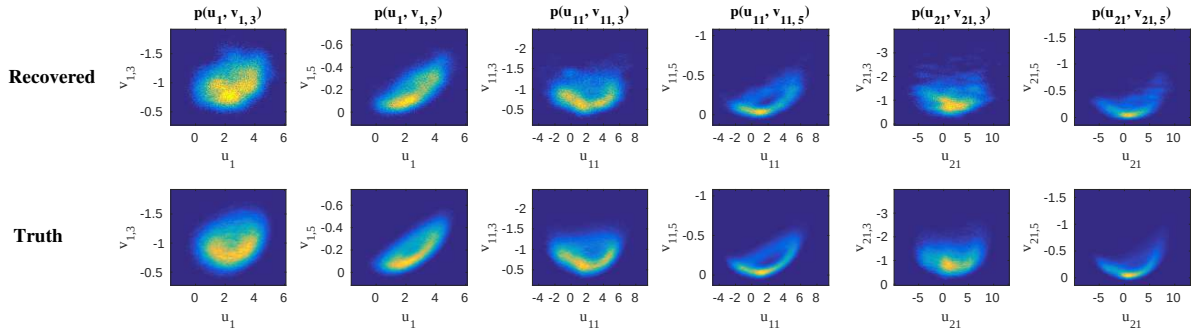

Fig. S19. [Two-layer inhomogeneous L-96 model]. 2D joint PDFs $p(u_i, v_{i,3})$ and $p(u_i, v_{i,5})$ at one fixed grid point i. Here the PDFs at $i = 1, 11$ and 21 are shown.

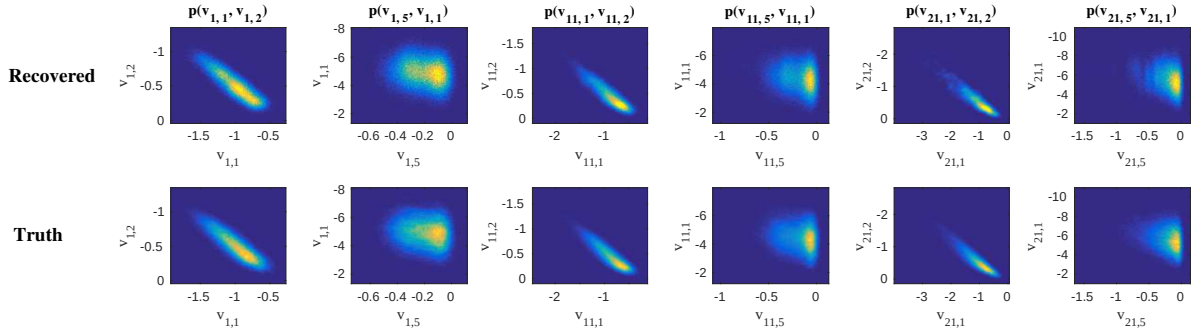

Fig. S20. [Two-layer inhomogeneous L-96 model]. 2D joint PDFs $p(v_{i,1}, v_{i,2})$ and $p(u_{i,5}, v_{i,1})$ at one fixed grid point i. Here the PDFs at $i = 1, 11$ and 21 are shown.

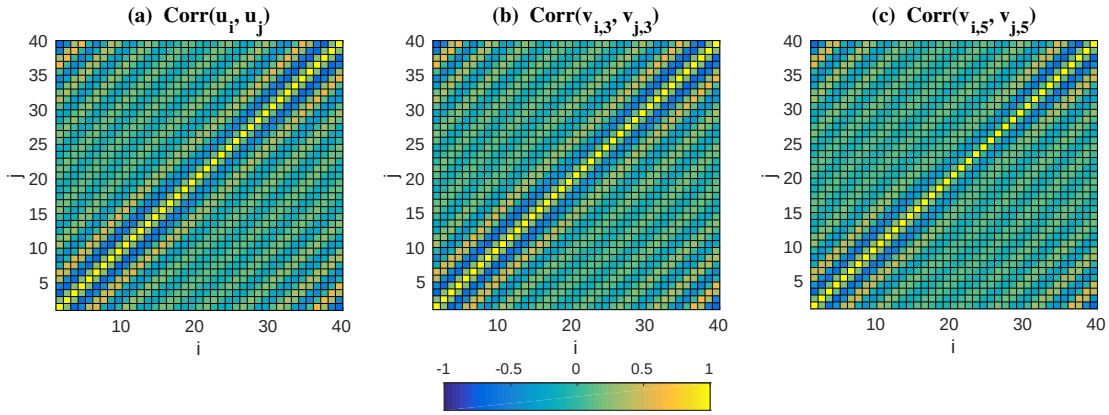

Fig. S21. [Two-layer inhomogeneous L-96 model]. Correlations between u_i and u_j (panel (a)), between $v_{i,3}$ and $v_{j,3}$ (panel (b)) and between $v_{i,5}$ and $v_{j,5}$ (panel (c)). The strongest negative correlation corresponds to $j = i \pm 2$ and the most significant positive correlation happens when $j = i \pm 5$.

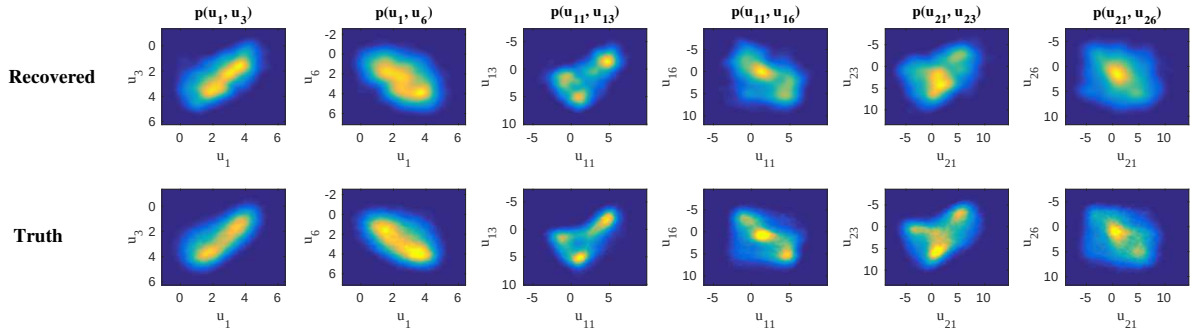

Fig. S22. [Two-layer inhomogeneous L-96 model]. Comparison of the truth and the recovered 2D PDFs of $p(u_i, u_{i+2})$ and $p(u_i, u_{i+5})$ with $i = 1, 11$ and 21. According to Fig. S21, $p(u_i, u_{i+2})$ and $p(u_i, u_{i+5})$ have the strongest negative and positive correlations, respectively, with each given i.

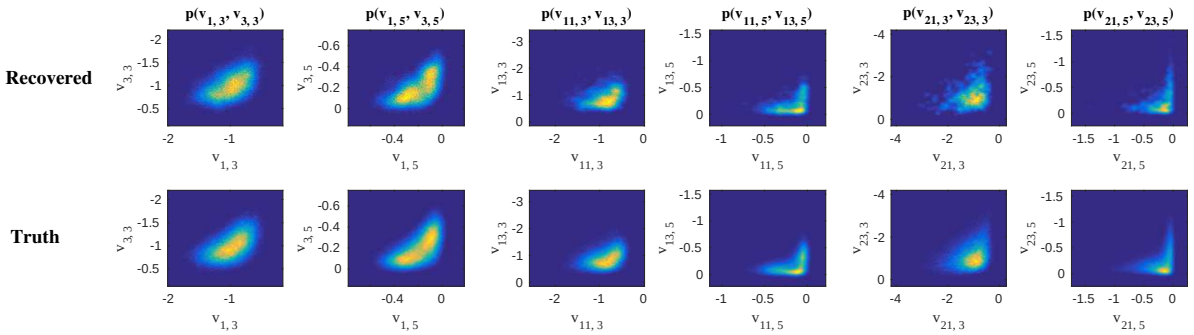

Fig. S23. [Two-layer inhomogeneous L-96 model]. Comparison of the truth and the recovered 2D PDFs of $p(v_{i,3}, v_{i+2,3})$ and $p(v_{i,5}, v_{i+2,5})$ with $i = 1, 11$ and 21. According to Fig. S21, $p(v_{i,3}, v_{i+2,3})$ and $p(v_{i,5}, v_{i+2,5})$ have the strongest negative correlations, respectively, with each given i.

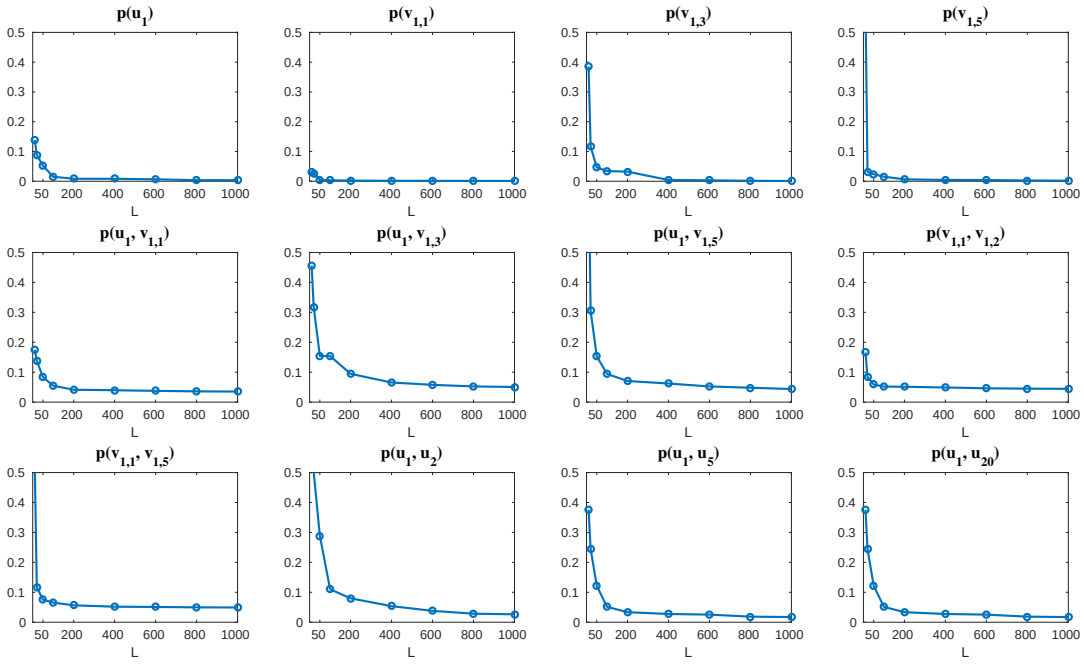

Fig. S24. [Two-layer inhomogeneous L-96 model]. Lack of information in the recovered PDFs compared with the truth at $i = 1$.

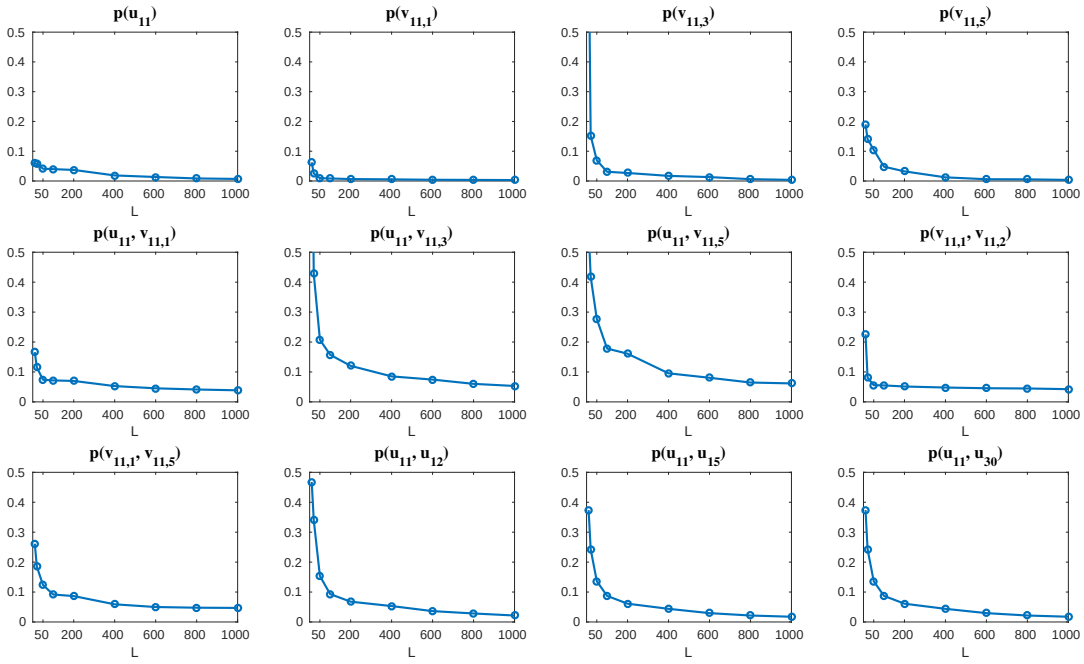

Fig. S25. [Two-layer inhomogeneous L-96 model]. Lack of information in the recovered PDFs compared with the truth at $i = 11$.

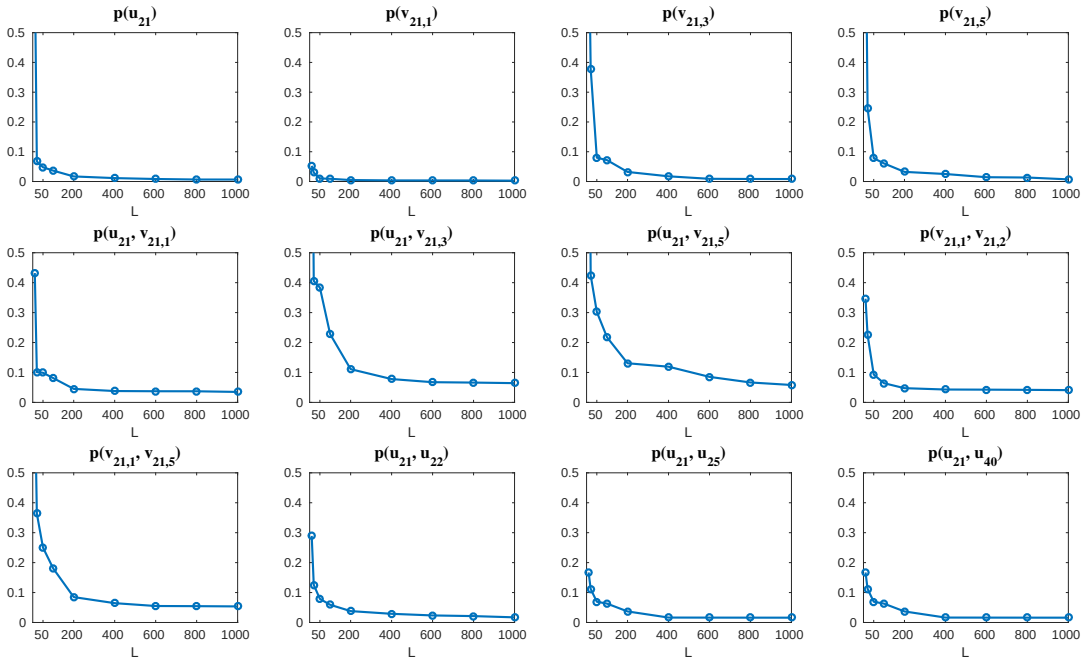

Fig. S26. [Two-layer inhomogeneous L-96 model]. Lack of information in the recovered PDFs compared with the truth at $i = 21$.