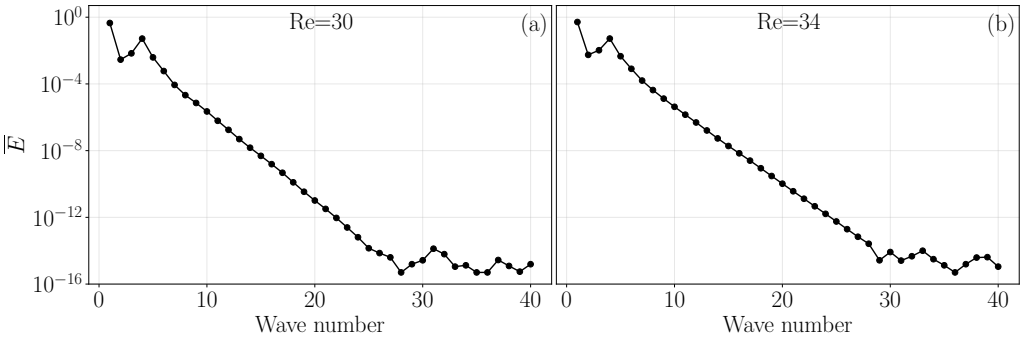

Figure 1: Energy spectra at Re=30, (a), and Re=34, (b).

Appendix S. Supplementary material

S.1. *Selection of number of Fourier modes*

Following Farazmand (2016), we select the number of Fourier modes by studying the convergence of the time-average, $\overline{(\cdot)}$, of the kinetic energy,

$$E = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi}\frac{1}{2}||\boldsymbol{u}||^2\,\mathrm{d}x_1\mathrm{d}x_2, \tag{S 1}$$

as a function of number of the fourier modes in the numerical solver. In both the quasiperiodic and chaotic cases, we select 32 wave numbers, Figure 1. For $k = 32$ wave numbers, we solve for $(2k+1)^2$ modes, $[-k, -k+1, \ldots, 0, \ldots, k-1, k]$ in each direction, where each mode consists of two complex components. Because the physical flowfield has real values, the coefficients of the negative wave numbers are the conjugates of the coefficient of the positive wave numbers, $c_{k1,k2} = c^*_{-k1,-k2}$. This results in $(2k+2)\times k\times 2\times 2 = 8448$ active degrees of freedom of the system ($c_{0,0}$ is constant as it represents the mean of the flow field) (chapter 6, Pope 2000).

S.2. *Jacobian-free computation of the Lyapunov exponents*

In this section, we describe the algorithm used to compute the first m largest Lyapunov exponents of the system. The algorithm requires the integration of the governing equations $m+1$ times, and does not require the computation of the Jacobian of the system. We consider a nonlinear autonomous dynamical system in the form of

$$\dot{\boldsymbol{q}} = \boldsymbol{f}(\boldsymbol{q}) \tag{S 2}$$

where $\boldsymbol{q}$ is the system's state and $\boldsymbol{f}$ is a nonlinear operator. In chaotic solutions, the norm of a perturbation $\boldsymbol{y}_i$, such that $\hat{\boldsymbol{q}}_i = \overline{\boldsymbol{q}} + \boldsymbol{y}_i$ with $||\boldsymbol{y}_i|| \ll ||\overline{\boldsymbol{q}}||$, grows in time until nonlinear saturation. For small enough times, $t_1 - t_0$, so that we avoid nonlinear saturation, the evolution of $\boldsymbol{y}_i$ can be computed as

$$\boldsymbol{y}_i(t_1) = \overline{\boldsymbol{q}}(t_1) - \hat{\boldsymbol{q}}_i(t_1), \tag{S 3}$$

where both elements in the right-hand side are computed by solving (S 2) with initial conditions equal to $\overline{\boldsymbol{q}}(t_0)$ and $\overline{\boldsymbol{q}}(t_0) + \boldsymbol{y}_i(t_0)$, respectively. The average exponential growth rate for the perturbation $\boldsymbol{y}_i$ between t_0 and t_1 is

$$\lambda = \frac{1}{t_1 - t_0}\ln\left(\frac{||\boldsymbol{y}(t_1)||}{||\boldsymbol{y}(t_0)||}\right), \tag{S 4}$$

where $||\cdot||$ indicates the L_2 norm. For $t_1 \to \infty$, almost all perturbations evolve with the same Λ_1, the dominant Lyapunov exponent

$$\Lambda_1 = \lim_{t_1 \to \infty} \frac{1}{t_1 - t_0} \ln\left(\frac{||\boldsymbol{y}(t_1)||}{||\boldsymbol{y}(t_0)||}\right), \tag{S 5}$$

as the component along the direction with maximum growth becomes dominant for sufficiently long times. However, due to saturation of the nonlinear equations (or instability of the linearized equations), computing Λ_1 is not straightforward.

To compute the growth along the m most unstable directions for long times, Benettin *et al.* (1980) proposed to periodically orthonormalize the evolution of the subspace spanned by m different perturbations. The algorithm works as follows. Every t_o, we orthonormalize the m perturbations and compute the future evolution of the orthonormalized basis:

$$\begin{aligned}
&\tilde{\boldsymbol{y}}_1(t) = \frac{\boldsymbol{y}_1(t)}{||\boldsymbol{y}_1(t)||} \quad \text{where} \quad \boldsymbol{y}_1(t - t_o) = \epsilon \tilde{\boldsymbol{y}}_1(t - t_o), \\
&\qquad \vdots \\
&\tilde{\boldsymbol{y}}_i(t) = \frac{\boldsymbol{y'}_i(t)}{||\boldsymbol{y'}_i(t)||}; \quad \boldsymbol{y'}_i = \boldsymbol{y}_i - \sum_{j=1}^{i-1} (\boldsymbol{y}_i^T \tilde{\boldsymbol{y}}_j)\tilde{\boldsymbol{y}}_j \quad \text{where} \quad \boldsymbol{y}_i(t - t_o) = \epsilon \tilde{\boldsymbol{y}}_i(t - t_o), \\
&\qquad \vdots \\
&\tilde{\boldsymbol{y}}_m(t) = \cdots
\end{aligned} \tag{S 6}$$

where $\epsilon \ll ||\boldsymbol{q}||$ is selected in order for initial condition to be infinitesimal and $\boldsymbol{y}_i(t)$ is computed using (S 3). For the first orthonormalization, the initial condition of the perturbations is random. At each orthonormalization, we store the average exponential growths, so that for the i-th direction at the k-th orthonormalization we have

$$\lambda_i^{(k)} = \frac{1}{t_o} \ln\left(\frac{||\boldsymbol{y}_i'(t)||}{||\boldsymbol{y}_i(t - t_o)||}\right) \tag{S 7}$$

where $||\boldsymbol{y}_i(t - t_o)|| = \epsilon$. After N_o orthonormalizations, the Lyapunov exponents are the average of the stored exponential growths

$$\Lambda_i = \frac{1}{N_o} \sum_{k=1}^{N_o} \lambda_i^{(k)}. \tag{S 8}$$

S.3. *Quasiperiodic dataset*

In this section, we provide additional analysis on the saturation of the NRMSE of the autoencoder in the quasiperiodic case, panel (a) in Fig. 4.

Figure 2 shows the NRMSE for the autoencoder (CNN) and different POD projections, which differ one from the other for the training and test set used. The CNN and POD errors are obtained by generating the subspace using the training set and then computing the NRMSE in the test set. The POD-train and POD-test are obtained by generating the reduced order model on the training and test sets, respectively, and then computing the NRMSE on the same set that was used to generate the reduced order model. We observe that both the CNN and POD show a saturation of the NRMSE, while the POD-test and POD-train do not. This indicates that there is a discrepancy between training data and the test data. We observed the same saturation behavior for different training sets of 30000 time units (results not shown), and we conclude that the saturation is due to the fact that a training set of 30000 time units

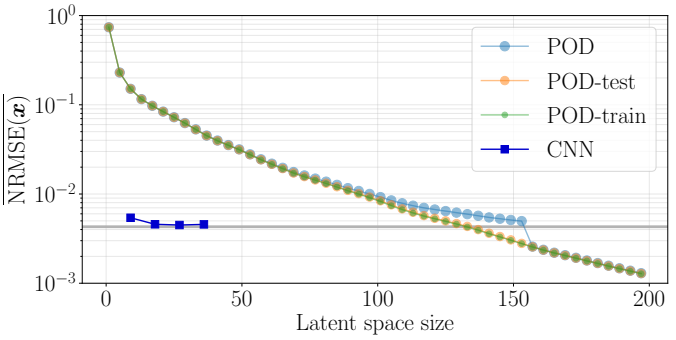

Figure 2: Discrepancy between training and test set in the quasiperiodic case.

is not fully representative of the entire trajectory of the system. This is a common situations in real world scenarios, where the available data is often incomplete.

S.4. *Extended results for the POD modes*

In Figure 3, we show the comparison of Figure 7 for the first ten POD modes.

S.5. *Dynamical content in the autoencoder loss function*

We provide an example of latent space tailored for the prediction in time of the system. We included in the autoencoder loss function the time-derivative of the system, which did not improve the results shown in §§5-6. We report the formulation here to (i) prevent fellow researchers from trying what we observed not to work and to (ii) provide a basis for the further development of dynamically-aware reduced-order modelling with machine learning in fluids. We minimise the error on the time derivative of the system to include information about the dynamical content of the system during the training of the autoencoder:

$$\mathcal{L} = \alpha_1 \sum_{i=1}^{N_t} \frac{1}{N_t} \frac{1}{N_{\text{phys}}} ||\hat{\boldsymbol{q}} - \boldsymbol{q}||^2 + \alpha_2 \sum_{i=1}^{N_t} \frac{1}{N_t} \frac{1}{N_{\text{phys}}} \left\| \frac{\mathrm{d}\hat{\boldsymbol{q}}}{\mathrm{d}t} - \frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t} \right\|^2, \tag{S 9}$$

where α_1 and α_2 are normalization coefficients selected for the two losses to have the same order of magnitude. They are equal to $1/\sigma(N_{\text{phys}}^{-1} ||\hat{\boldsymbol{q}} - \boldsymbol{q}||^2)$ and $1/\sigma(N_{\text{phys}}^{-1} \left\| \frac{d\hat{\boldsymbol{q}}}{dt} - \frac{d\boldsymbol{q}}{dt} \right\|^2)$, respectively; $\sigma(\cdot)$ is the standard deviation computed over the training set. In this work, the time-derivative is computed using by first-order finite difference:

$$\frac{\mathrm{d}\boldsymbol{q}}{\mathrm{d}t}(t_i) = \frac{\boldsymbol{q}(t_i + \delta t) - \boldsymbol{q}(t_i)}{\delta t}, \tag{S 10}$$

where δt is defined in §2. Other possible choices are automatic differentiation (Racca & Magri 2021*a*) or the governing equations (if known). The purpose of minimising the error on the time-derivative is for the autoencoder to reconstruct not only the structures that account for the majority of the energy, but also the structures that contribute to the change in time of the state of the system.

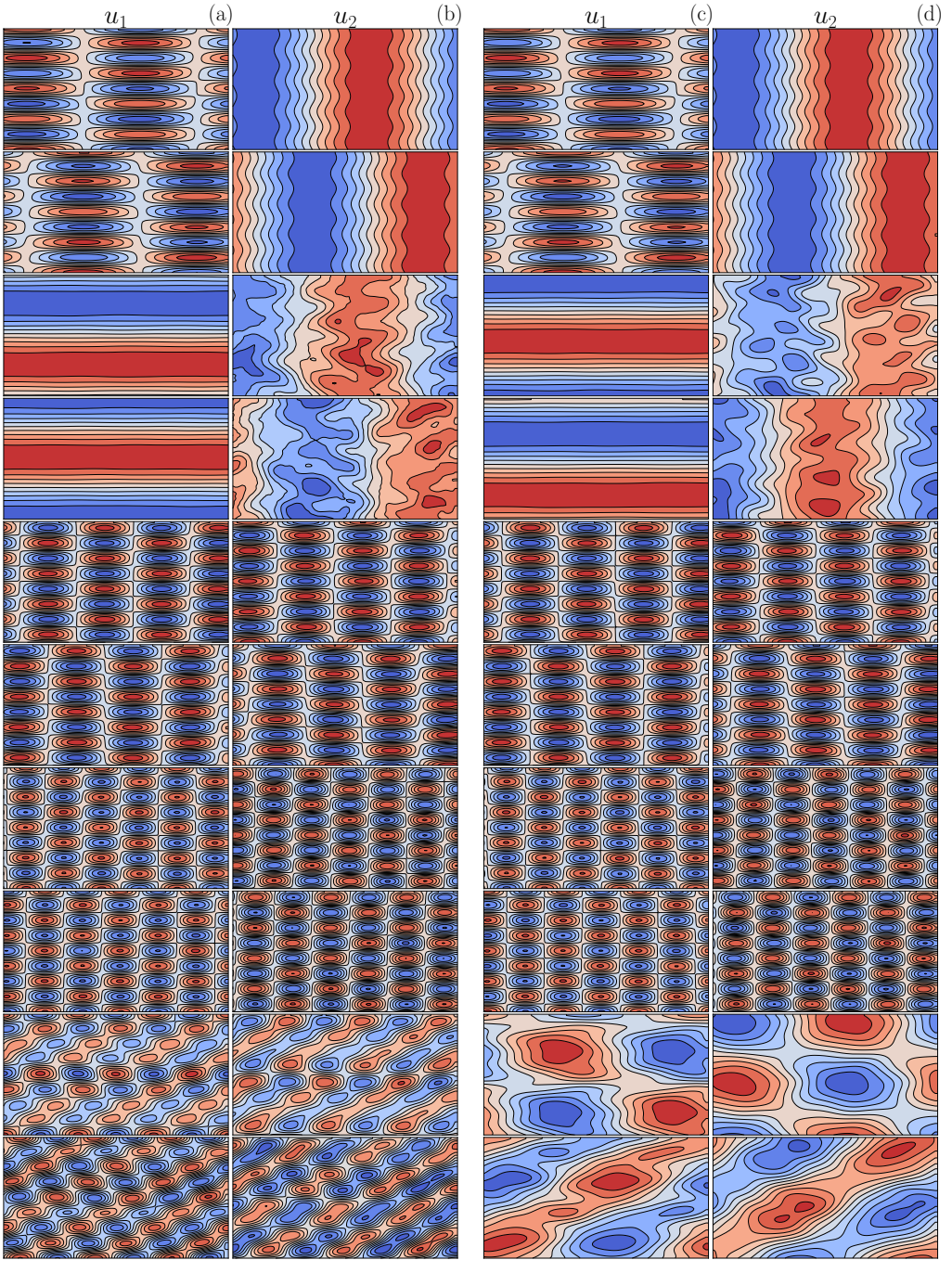

Figure 3: First ten POD modes, one per row, for the reconstructed flow field based on 4 POD modes in latent space (a)-(b) and the true flow field (c)-(d). POD modes are sorted in descending order with first row being the first mode.