PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Schmidt OT. 2026 Data-driven forecasting of high-dimensional transient and stationary processes via space–time projection. *Proc. R. Soc. A* **482**: 20250454.
https://doi.org/10.1098/rspa.2025.0454

Received: 28 May 2025
Accepted: 17 November 2025

Subject Areas:
fluid mechanics, mathematical modelling, applied mathematics

Keywords:
forecasting, model-order reduction, proper orthogonal decomposition, Hankel

Author for correspondence:
Oliver T. Schmidt
e-mail: oschmidt@ucsd.edu

THE ROYAL SOCIETY
PUBLISHING

Data-driven forecasting of high-dimensional transient and stationary processes via space–time projection

Oliver T. Schmidt

Department of Mechanical and Aerospace Engineering,
University of California San Diego, La Jolla, CA 92093, USA

OTS, 0000-0002-7097-0235

Space–time projection (STP) is introduced as a data-driven forecasting approach for high-dimensional, time-resolved data. The method computes extended space–time proper orthogonal modes from training data spanning a prediction horizon comprising both hindcast and forecast intervals. Forecasts are generated by projecting the hindcast portion of these modes onto new data, leveraging their orthogonality and optimal correlation with the forecast extension. Rooted in proper orthogonal decomposition (POD) theory, dimensionality reduction and time-delay embedding are intrinsic to the approach. The only tunable parameters are the truncation rank and the hindcast length; no additional hyperparameters are required. Hindcast accuracy serves as a reliable indicator for short-term forecast accuracy. The method's efficacy is demonstrated using two datasets: transient, highly anisotropic simulations of supernova explosions in a turbulent interstellar medium and experimental velocity fields of a turbulent high-subsonic engineering flow. In a comparative study with standard dynamic mode decomposition (DMD) and long short-term memory (LSTM) networks (acknowledging that alternative architectures or training strategies may yield different outcomes) STP achieved the lowest errors at short and long lead times and was comparable at intermediate horizons. Considering its simplicity and robust performance, STP offers an interpretable and competitive baseline for forecasting high-dimensional transient chaotic processes, relying purely on spatio-temporal correlation.

1. Introduction

Forecasting the evolution of high-dimensional real-world processes is a formidable challenge in a variety of fields ranging from atmospheric science to engineering and astrophysics. Such systems frequently exhibit nonlinear dynamics and chaos, imposing fundamental limits on predictability. When available and computationally tractable, physics-based numerical models, typically in the form of discretized systems of partial differential equations representing conservation laws for mass, momentum, and energy, offer powerful predictive capabilities. These numerical models are especially effective when combined with data assimilation techniques [1], such as Kalman filtering [2], 4D-Var [3] and ensemble-based methods [4], to incorporate measurement data. Arguably the most notable example is numerical weather prediction [5].

This work addresses the complementary challenge of forecasting large spatio-temporal datasets through purely data-driven methods, regardless of their origin or the availability of governing physical models. This class of problems is often referred to as multivariate or multichannel time-series forecasting [6,7]. Widely known methods for forecasting such data include autoregressive integrated moving average, vector autoregression, Kalman filtering, state-space models [8] and various extensions of singular spectrum analysis [9]. The choice of an appropriate forecasting method typically depends on the statistical nature of the data; transient, stationary and cyclo-stationary processes are ubiquitous across many fields, each requiring different treatments or preprocessing techniques.

Many of these forecasting methods rely fundamentally on matrix decompositions, most commonly singular value decomposition (SVD) applied directly to data, Hankel or correlation matrices. As a result, these methods often directly make use of proper orthogonal decomposition (POD), or indirectly leverage the same mathematical properties of optimality and orthogonality provided by SVD. While the original formulation of POD [10,11] defines modes that are dependent on both space and time, the most commonly used variant, referred to simply as POD, separates the flow data into spatial modes and time-dependent expansion coefficients [12,13]. To distinguish clearly between these different formulations, we refer to this widely used version as space-only or method-of-snapshots POD. Although the theoretical foundations of the most general form of space–time POD have long been established, algorithmic implementations and practical applications remain extremely rare, with [14,15] being exceptions. Only recently, the equivalence between space–time POD and the SVD of a Hankel matrix was explicitly demonstrated by the authors of [16], who also showed that space-only POD and spectral POD (SPOD; [17,18]) arise naturally as the limiting cases of space–time POD in the short- and long-time limits, respectively.

Extended POD [19], a key enabler of this work, was initially proposed as a method to analyse correlated events in turbulent flows. It was first applied to investigate correlations between spatially distinct regions, and subsequently generalized and rigorously formalized by Boree [20]. Similar to POD, dynamic mode decomposition (DMD; [21,22]) is fundamentally a modal decomposition technique [23]. However, grounded in the Koopman theory [24], DMD yields a discrete propagation operator for the system state along with modes characterized by complex frequencies. This property enables direct forecasting by propagating the modes forward in time using their intrinsic oscillation frequencies and corresponding growth or decay rates. Similarly, SPOD yields modes associated with distinct frequencies, making them inherently suitable for forecasting without further modelling. That said, without additional modelling or rank truncation, SPOD forecasts reduce to repetitions of previously observed data owing to the inherent stationarity assumption. A stochastic model combining SPOD for dimensionality reduction and Koopman theory for forecasting was recently proposed by Chu & Schmidt [25].

Machine learning (ML) methods represent another prominent category of forecasting techniques. Models specifically designed or readily applicable to time-series forecasting include recurrent neural networks [26], long short-term memory networks (LSTM; [27]) and simplified gated recurrent units [28]. Additional ML-based approaches include temporal convolutional networks [29], transformer-based architectures [30] and reservoir computing [31,32]. Two contributions explicitly targeting high-dimensional physical system forecasting are [33], using LSTM, and [34], demonstrating reservoir computing. While these studies successfully forecast chaotic model systems such as the Lorenz system and the Kuramoto–Sivashinsky equation, they typically cannot be applied directly to large datasets arising from high-fidelity simulations or experimental measurements, such as particle image velocimetry (PIV; [35]). For such data, dimensionality reduction is usually required first, making ML-approaches sequential. One example, used later in this paper for comparison with the proposed space–time projection (STP) method, incorporates POD for dimensionality reduction, followed by ML-based forecasting of the resulting POD coefficients representing latent temporal dynamics. This approach was explored in [36] for both POD and SPOD.

This paper introduces STP as a forecasting method that exclusively relies on correlation information to make forecasts of high-dimensional fields. We begin by presenting the method formulation in §2, accompanied by a visual summary in figure 1 and concluded with a summary of the algorithm in §2a. In the results section, §3, we demonstrate the method's capabilities using two large datasets, both exhibiting highly chaotic dynamics, but one numerical and transient, the other experimental and stationary. First, in §3a, we examine the transient, non-stationary example: a supernova explosion evolving within a turbulent interstellar medium, simulated by Hirashima *et al.* [37]. Second, in §3b, we examine a statistically stationary case using time-resolved velocity field measurements from a Mach 0.6 turbulent cavity flow experiment by [38], a canonical flow configuration representative of landing gear bays and cavities formed by high-lift devices on aircraft [39]. In §3c, two important aspects of STP's forecasting performance are evaluated: its sensitivity of forecast accuracy to the available training data size, and its forecast accuracy in direct comparison to LSTM regression neural networks trained on POD-reduced coefficients and classical DMD. Finally, in §4, we summarize the proposed approach, comment on the mathematical versus physical interpretation of the modes that form the basis, and outline future directions and potential applications.

2. Methods

Given n snapshots, $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, spanning the hindcast horizon, our goal is to predict the next m snapshots, $\mathbf{u}_{n+1}, \mathbf{u}_{n+2}, \ldots, \mathbf{u}_{n+m}$, over the forecast horizon. We refer to the combined hindcast and forecast horizons as the prediction horizon. The proposed method is based on an ensemble of k realizations of the transient stochastic process we wish to predict, with each realization sampled over the prediction horizon. We denote the snapshot representing the system state at time t_i ($i = 1, 2, \ldots, n+m$) in the jth realization ($j = 1, 2, \ldots, k$) as

$$\mathbf{u}_i^{[j]} = \mathbf{u}^{[j]}(\mathbf{x}, t_i), \tag{2.1}$$

stored as a flattened vector comprising p degrees of freedom (typically, the total number of grid points multiplied by the number of variables). The entire sequence of $n+m$ snapshots for the jth realization constitutes an episode, describing a single trajectory of the stochastic process. Alternatively, we compactly write the ensemble of k episodes as $\{\mathbf{u}_i^{[j]}\}_{i=1}^{n+m}$, where i indexes snapshots within an episode, and j indexes episodes. We assume that the ensemble mean,

$$\bar{\mathbf{u}}_i = \frac{1}{k}\sum_{j=1}^{k} \mathbf{u}_i^{[j]}, \tag{2.2}$$

computed over k trajectories for each time instance i, is already subtracted from the data. This preprocessing step allows the eigenvalues to be interpreted as variances or, depending on the choice of variables and weighting, as mode energies. We construct the $np \times 1$ hindcast data vector $\mathbf{q}_- = [\mathbf{u}_1^{\mathrm{T}}\ \ \mathbf{u}_2^{\mathrm{T}} \cdots \mathbf{u}_n^{\mathrm{T}}]^{\mathrm{T}}$ by reshaping each snapshot $\mathbf{u}_i$ into a one-dimensional vector and then stacking them sequentially to form a single column vector. The subscript $(\cdot)_-$ denotes vectors and matrices spanning the hindcast horizon, $(\cdot)_+$ the forecast horizon and $(\cdot)_\pm$ the entire prediction horizon, encapsulating both the hindcast and forecast.

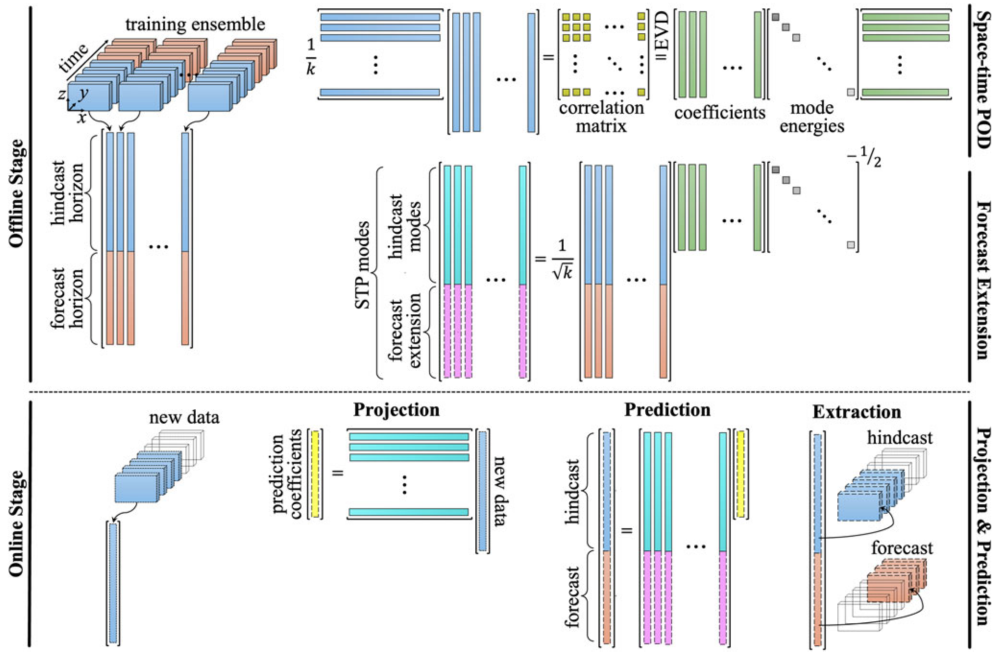

Figure 1. Schematic of the STP algorithm. Step 1: construct the hindcast data matrix and solve the ensemble space–time POD eigenvalue problem, equation (2.6), to obtain the hindcast expansion coefficients. Step 2: compute the STP (extended hindcast) modes using equation (2.10). Step 3: approximate the hindcast expansion coefficients via projection, equation (2.11), and expand the predicted trajectory using equation (2.12).

Given an ensemble of k realizations of the stochastic process, we construct the $np \times k$ hindcast data matrix,

$$\mathbf{Q}_- = [\mathbf{q}_-^{[1]} \quad \mathbf{q}_-^{[2]} \quad \cdots \quad \mathbf{q}_-^{[k]}] = \begin{bmatrix} \mathbf{u}_1^{[1]} & \mathbf{u}_1^{[2]} & \cdots & \mathbf{u}_1^{[k]} \\ \mathbf{u}_2^{[1]} & \mathbf{u}_2^{[2]} & \cdots & \mathbf{u}_2^{[k]} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_n^{[1]} & \mathbf{u}_n^{[2]} & \cdots & \mathbf{u}_n^{[k]} \end{bmatrix}, \tag{2.3}$$

and the $(n+m)p \times k$ prediction data matrix,

$$\mathbf{Q}_{\pm} = \begin{bmatrix} \mathbf{q}_-^{[1]} & \mathbf{q}_-^{[2]} & \cdots & \mathbf{q}_-^{[k]} \\ \mathbf{q}_+^{[1]} & \mathbf{q}_+^{[2]} & \cdots & \mathbf{q}_+^{[k]} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^{[1]} & \mathbf{u}_1^{[2]} & \cdots & \mathbf{u}_1^{[k]} \\ \mathbf{u}_2^{[1]} & \mathbf{u}_2^{[2]} & \cdots & \mathbf{u}_2^{[k]} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_n^{[1]} & \mathbf{u}_n^{[2]} & \cdots & \mathbf{u}_n^{[k]} \\ \mathbf{u}_{n+1}^{[1]} & \mathbf{u}_{n+1}^{[2]} & \cdots & \mathbf{u}_{n+1}^{[k]} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_{n+m}^{[1]} & \mathbf{u}_{n+m}^{[2]} & \cdots & \mathbf{u}_{n+m}^{[k]} \end{bmatrix}, \tag{2.4}$$

which includes full episodes spanning both the hindcast and forecast horizons. Analogous to the hindcast data, the forecast data are flattened into the $mp \times 1$ vector $\mathbf{q}_+ = [\mathbf{u}_{n+1}^{\mathrm{T}} \quad \mathbf{u}_{n+2}^{\mathrm{T}} \cdots \mathbf{u}_{n+m}^{\mathrm{T}}]^{\mathrm{T}}$.

At the core of the method is the ensemble space–time POD of the hindcast data matrix $\mathbf{Q}_-$. Like other POD variants, the space–time POD problem can be solved using eigenvalue decomposition of either the spatio-temporal or ensemble correlation matrix, or directly through the SVD of the data matrix. Since the ensemble size k is typically much smaller than the total number of spatio-temporal degrees of freedom, np, we perform the computation using the Hermitian $k \times k$ ensemble sample correlation matrix,

$$\mathbf{C}_- = \frac{1}{k}\mathbf{Q}_-^H \mathbf{W} \mathbf{Q}_-, \tag{2.5}$$

and compute its eigenvalue decomposition,

$$\mathbf{C}_- \boldsymbol{\Psi}_- = \boldsymbol{\Psi}_- \boldsymbol{\Lambda}, \tag{2.6}$$

to obtain the ensemble coefficient matrix, scaled such that $\boldsymbol{\Psi}_-^H \boldsymbol{\Psi}_- = \mathbf{I}$. For generality, the diagonal weight matrix $\mathbf{W} \succeq 0$ is introduced in equation (2.5) to account for spatial and variable weighting. The diagonal eigenvalue matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ contains the mode energies, ranked by their magnitude. Note that, depending on the choice of variables, these eigenvalues may not strictly represent physical energy; rather, they quantify the captured variance of each mode. In any case, they indicate the relative importance of individual modes and can always be interpreted as the squared amplitudes of the modes.

If rank reduction is desired, whether to suppress noise by discarding low-energy components or to reduce computational complexity, we can truncate the representation to retain only the leading $r < k$ components. Specifically, we truncate the columns of the coefficient matrix $\boldsymbol{\Psi}_-$ and reduce the eigenvalue matrix to $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r)$. With this reduced representation, all subsequent steps (mode reconstruction, projection and forecasting) are carried out using the truncated basis. We omit explicit notation for this rank truncation to keep the presentation concise. First, the *hindcast modes* are obtained by expanding the hindcast data using the ensemble coefficients as

$$\boldsymbol{\Phi}_- = \frac{1}{\sqrt{k}} \mathbf{Q}_- \boldsymbol{\Psi}_- \boldsymbol{\Lambda}^{-1/2} \quad \text{(hindcast modes)}, \tag{2.7}$$

where the scaling ensures that the hindcast modes are orthonormal with $\boldsymbol{\Phi}_-^H \mathbf{W} \boldsymbol{\Phi}_- = \mathbf{I}$. Since the modal basis $\boldsymbol{\Phi}_-$ spans the same subspace as the data $\mathbf{Q}_-$, we can express the latter as a linear combination of the modes,

$$\mathbf{Q}_- = \boldsymbol{\Phi}_- \mathbf{A}_-. \tag{2.8}$$

The matrix of expansion coefficients $\mathbf{A}_-$ can be related to $\boldsymbol{\Psi}_-$ via $\mathbf{A}_- = \sqrt{k}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Psi}_-^H$, or, alternatively, obtained by projecting the data directly onto the modes as

$$\mathbf{A}_- = \boldsymbol{\Phi}_-^H \mathbf{W} \mathbf{Q}_-, \tag{2.9}$$

which has the orthogonality property $(1/k)\mathbf{A}_- \mathbf{A}_-^H = \boldsymbol{\Lambda}$. Together with the orthonormality of the modes, this property makes clear that we can interpret the square root of the eigenvalues as mode amplitudes.

To make predictions over the forecast horizon, we define the *STP modes* as the extended hindcast modes,

$$\boldsymbol{\Phi}_\pm^\star = \frac{1}{\sqrt{k}} \mathbf{Q}_\pm \boldsymbol{\Psi}_- \boldsymbol{\Lambda}^{-1/2} \quad \text{(STP modes)}, \tag{2.10}$$

obtained analogously to equation (2.7) by expanding the ensemble data of the full trajectories into extended modes, $\boldsymbol{\Phi}_\pm^\star$, which span both the hindcast and forecast horizons. This is done by replacing $\mathbf{Q}_-$ in equation (2.7) with $\mathbf{Q}_\pm$, while keeping the same hindcast expansion coefficients. We use the superscript $(\cdot)^\star$ to denote prediction quantities that are analogous to but deviate from their exact counterparts. Note that $\boldsymbol{\Phi}_\pm^\star$ are not the space–time modes $\boldsymbol{\Phi}_\pm$ related to $\mathbf{C}_\pm$, but rather extended hindcast modes. In the expansion equation (2.10), the hindcast components remain unchanged, meaning the first np elements of each STP mode $\boldsymbol{\phi}_\pm^\star$ (the columns of $\boldsymbol{\Phi}_\pm^\star$) exactly match the corresponding hindcast mode $\boldsymbol{\phi}_-$. However, these modes are extended over the forecast horizon by an additional $mp \times 1$ vector, $\boldsymbol{\phi}_{+}^{\star}$, which retains the portion, and only that, of the forecast data that is correlated with the hindcast. This key correlation property of extended POD was elegantly proven by Boree [20] and directly applies to the extended space–time, or STP modes $\boldsymbol{\phi}_{\pm}^{\star} = [\boldsymbol{\phi}_{-}^{\mathrm{T}} \boldsymbol{\phi}_{+}^{\star T}]^{\mathrm{T}}$. We refer to proposition 1 in [20] for a formal proof. An equivalent way to compute the extended hindcast modes is to perform the eigenvalue decomposition of $\mathbf{C}\pm$ constructed from $\mathbf{Q}\pm$, using an extended block-diagonal weight matrix with zero weights on the extension.

Given a new hindcast trajectory, $\mathbf{q}_{-}^{\mathrm{new}}$, consisting of n stacked snapshots, our goal is to forecast the next m time instances of the flow, $\mathbf{q}_{+}^{\star}$. To compute the forecast using an expansion in the STP mode basis, $\boldsymbol{\Phi}_{\pm}$, we need the corresponding expansion coefficients, $\mathbf{a}_{\pm}^{\star}$, which are unknown. The key idea is to approximate $\mathbf{a}_{\pm}^{\star}$ using $\mathbf{a}_{-}^{\star}$, the hindcast expansion coefficients and leverage the correlation property of extended POD to ensure an optimal forecast. The hindcast expansion coefficients $\mathbf{a}_{-}^{\star}$ are obtained in accordance with equation (2.9) by projecting the new hindcast data onto the precomputed hindcast basis,

$$\mathbf{a}_{-}^{\star} = \boldsymbol{\Phi}_{-}^{H} \mathbf{W} \mathbf{q}_{-}^{\mathrm{new}}. \tag{2.11}$$

Under the approximation that $\mathbf{a}_{\pm}^{\star} \approx \mathbf{a}_{-}^{\star}$, we can readily reconstruct the predicted full trajectory, $\mathbf{q}_{\pm}^{\star} \approx \mathbf{q}_{\pm}$, as

$$\mathbf{q}_{\pm}^{\star} = \boldsymbol{\Phi}_{\pm}^{\star} \mathbf{a}_{-}^{\star}. \tag{2.12}$$

That is, by expanding the STP mode basis using the hindcast coefficients, see equation (2.8). The prediction, $\mathbf{q}_{\pm}^{\star} = [\mathbf{q}_{-}^{\star T} \ \mathbf{q}_{+}^{\star T}]^{\mathrm{T}}$, can then be separated into the hindcast, $\mathbf{q}_{-}^{\star}$, and the forecast, $\mathbf{q}_{+}^{\star}$. Alternatively, we can separate equation (2.12) and compute the hindcast and forecast components separately as $\mathbf{q}_{-}^{\star} = \boldsymbol{\Phi}_{-}^{\star} \mathbf{a}_{-}^{\star}$ and $\mathbf{q}_{+}^{\star} = \boldsymbol{\Phi}_{+}^{\star} \mathbf{a}_{-}^{\star}$, respectively. A difference between the new data, $\mathbf{q}_{-}^{\mathrm{new}}$, and the hindcast, $\mathbf{q}_{-}^{\star}$, is expected for real-world stochastic physical processes, where the new data may not lie within the span of the training data used to compute the STP basis via equation (2.6). As the examples demonstrate, this hindcast error provides valuable insight into the accuracy of the forecast. The key steps of the algorithm are visually summarized in figure 1. For simplicity, uniform weighting with $\mathbf{W} = \mathbf{I}$ is assumed for figure 1 and the definitions of the following error metrics.

We use the root-mean-squared (RMS) error to quantify the difference between the true state, $\mathbf{u}_i = \mathbf{u}(\mathbf{x}, t_i)$, and the predicted fields, $\mathbf{u}_i^{\star} = \mathbf{u}^{\star}(\mathbf{x}, t_i)$, at the ith time instant. The RMS error for a single trajectory is hence defined as

$$e_{\mathrm{RMS},i} = \frac{1}{\sqrt{p}} ||\mathbf{u}_i - \mathbf{u}_i^{\star}||_2, \tag{2.13}$$

the mean RMS error over an ensemble of k trajectories by

$$\bar{e}_{\mathrm{RMS},i} = \frac{1}{k} \sum_{j=1}^{k} e_{\mathrm{RMS},i}^{[j]} \tag{2.14}$$

and the corresponding standard deviation, which quantifies the spread of the error, is calculated as

$$\sigma_{e_{\mathrm{RMS},i}} = \sqrt{\frac{1}{k-1} \sum_{j=1}^{k} (e_{\mathrm{RMS},i}^{[j]} - \bar{e}_{\mathrm{RMS},i})^2}. \tag{2.15}$$

Together, these metrics provide a comprehensive assessment of the prediction accuracy, both in terms of average performance and variability across trajectories.

(a) STP algorithm

Finally, we present the STP algorithm corresponding to equations (2.3)–(2.12) in a procedural form, explicitly separating the offline (training) and online (deployment or inference) stages.

Algorithm 1: Space–time projection.

```
Offline (once)
    Input  : mean-subtracted training episodes $\{\mathbf{u}_i^{[j]}\}_{i=1}^{n+m}$ for $j = 1, \dots, k$; weight $\mathbf{W} \succeq 0$
             (default $\mathbf{I}$); hindcast length $n$, forecast length $m$, (optional) truncation rank $r$.
    Output: hindcast modes $\boldsymbol{\Phi}_- \in \mathbb{C}^{(np)\times r}$ and extended modes $\boldsymbol{\Phi}_\pm^\star \in \mathbb{C}^{((n+m)p)\times r}$.
    $\mathbf{q}_-^{[j]} \leftarrow [\mathbf{u}_1^{[j]T} \cdots \mathbf{u}_n^{[j]T}]^T$,   $\mathbf{q}_\pm^{[j]} \leftarrow [\mathbf{u}_1^{[j]T} \cdots \mathbf{u}_{n+m}^{[j]T}]^{\mathrm{T}}$          // assemble data
    $\mathbf{Q}_- \leftarrow [\mathbf{q}_-^{[1]} \cdots \mathbf{q}_-^{[k]}]$,   $\mathbf{Q}_\pm \leftarrow [\mathbf{q}_\pm^{[1]} \cdots \mathbf{q}_\pm^{[k]}]$
    $\mathbf{C}_- \leftarrow \frac{1}{k}\mathbf{Q}_-^H \mathbf{W} \mathbf{Q}_-$;   compute $\mathbf{C}_- \boldsymbol{\Psi}_- = \boldsymbol{\Psi}_- \Lambda$ with $\boldsymbol{\Psi}_-^H \boldsymbol{\Psi}_- = \mathbf{I}$  // hindcast ST-POD
    (Optional) compute economy eigendecomposition, retaining only $r$ leading modes.
    $\boldsymbol{\Phi}_- \leftarrow \frac{1}{\sqrt{k}} \mathbf{Q}_- \boldsymbol{\Psi}_- \Lambda^{-1/2}$          // hindcast modes
    $\boldsymbol{\Phi}_\pm^\star \leftarrow \frac{1}{\sqrt{k}} \mathbf{Q}_\pm \boldsymbol{\Psi}_- \Lambda^{-1/2}$; partition $\boldsymbol{\Phi}_\pm^\star = \begin{bmatrix} \boldsymbol{\Phi}_-^\star \\ \boldsymbol{\Phi}_+^\star \end{bmatrix}$ with $\boldsymbol{\Phi}_-^\star = \boldsymbol{\Phi}_-$   // extended modes
Online (per query)
    Input  : new hindcast data $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$; precomputed $\boldsymbol{\Phi}_-$ and $\boldsymbol{\Phi}_\pm^\star$.
    Output: forecast $\{\mathbf{u}_{n+1}^\star, \dots, \mathbf{u}_{n+m}^\star\}$.
    $\mathbf{q}_- \leftarrow [\mathbf{u}_1^{\mathrm{T}} \cdots \mathbf{u}_n^{\mathrm{T}}]^{\mathrm{T}}$;   $\mathbf{a}_-^\star \leftarrow \boldsymbol{\Phi}_-^H \mathbf{W} \mathbf{q}_-$          // project hindcast
    $\mathbf{q}_\pm^\star \leftarrow \boldsymbol{\Phi}_\pm^\star \mathbf{a}_-^\star$;   extract $\mathbf{q}_+^\star$ and unstack to $\{\mathbf{u}_{n+1}^\star, \dots, \mathbf{u}_{n+m}^\star\}$          // prediction
```

The offline stage constructs the space–time basis from the ensemble data, while the online stage projects new hindcast data onto this basis to produce the forecast.

3. Results

STP is demonstrated on two large, distinct datasets, each presenting unique challenges. The first example, discussed in §3a, involves numerical data of a supernova explosion [37] with a highly anisotropic shell rapidly expanding into a turbulent interstellar medium, whose localized and transient features make it particularly challenging. The second example, examined in §3b, considers experimental velocity field measurements of the flow over and inside an open cavity at high speeds [38], where the main difficulties arise from the broadband nature of the stationary turbulent flow and the convective behaviour of its coherent structures. Figures 2 and 3 introduce the two cases and highlight these challenges. Since both datasets are sampled on equidistant grids and we are only concerned with relative comparisons, we use uniform weighting with $\mathbf{W} = \mathbf{I}$.

(a) Supernova explosion

The first example considers a supernova explosion in a turbulent interstellar medium, where a blast wave propagates through a dense, cool gas cloud, generating complex and anisotropic shell structures. The simulations, conducted by Hirashima *et al.* [37] using the smoothed particle hydrodynamics code ASURA-FDPS [40,41], were obtained from the The Well data repository [42]. The set-up models a supernova, the explosion of a massive star, within a compressible monatomic ideal gas, governed by the equation of state with a specific heat ratio $\gamma = 5/3$. The simulated gas represents the interstellar medium of the Milky Way galaxy with an initial gas-particle mass of one solar mass. To initiate the explosion, a thermal energy of 10^{51} erg is deposited at the domain centre, rapidly heating the gas to temperatures around 10^7 K and generating a blast wave. The initial conditions feature 820 randomly seeded molecular clouds with turbulent statistics following Burgers' turbulence model.

For this demonstration, we use the temperature field $T(\mathbf{x}, t)$ as the sole variable. For each snapshot, the ensemble mean, calculated using equation (2.2), is subtracted and we define the

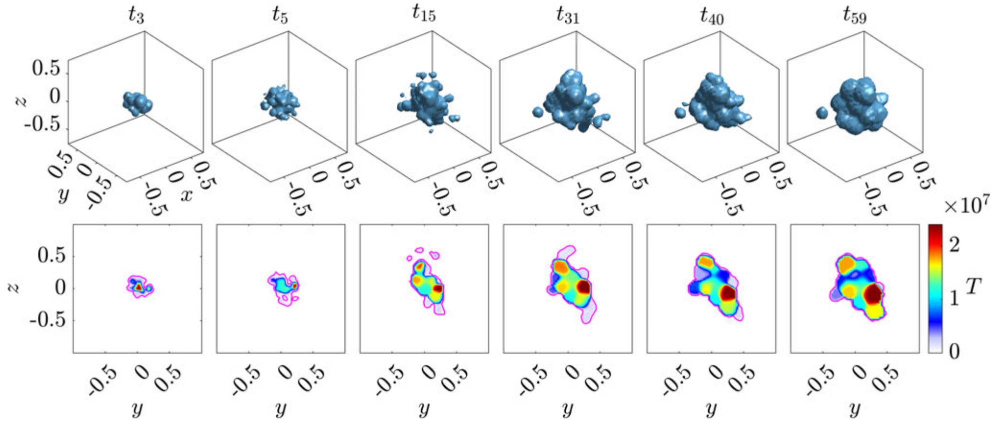

Figure 2. Overview of the numerical supernova simulation. The top row shows instantaneous isocontours of temperature at 1% of its maximum value, highlighting the expanding supernova shell at six representative time instances. The bottom row presents the corresponding temperature fields in the y–z plane at $x = 0$. The magenta contour line indicates the isovalue used in the three-dimensional visualization of the shell. The first realization from the training dataset is shown as an example. Coordinates are non-dimensionalized by the domain length of 60 pc.

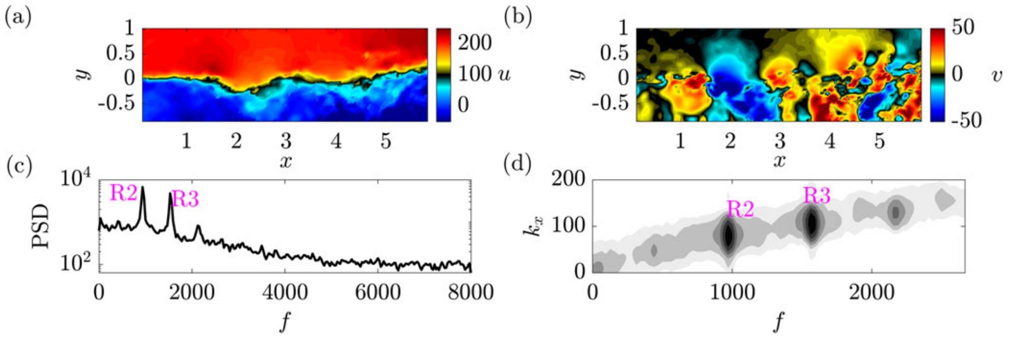

Figure 3. Overview of experimental cavity flow. (a) Instantaneous streamwise velocity u; (b) instantaneous normal velocity v; (c) power spectrum of v at $(x, y) = (5, 0)$ and (d) frequency–wavenumber diagram along $(x, y = 0)$. The peaks identified in (c) and (d) correspond to the two dominant Rossiter tones. The incoming flow travels over the cavity ($y > 0$) at Mach 0.6 and recirculates within the cavity walls ($y < 0$) while undergoing violent oscillations. Coordinates are non-dimensionalized by the cavity depth D, while all other quantities are reported in SI units.

state vector for the ith snapshot as

$$\mathbf{u}_i = T'_i,$$

where T'_i is the flattened vector of the fluctuating temperature. The repository provides temperature data at 59 time steps on a uniform Cartesian $64 \times 64 \times 64$ grid, spanning 60 pc (parsecs; 1 pc $\approx 3.086 \times 10^{16}$ m $\approx$ 3.26 light-years) in each direction. For this single scalar variable, temperature, each snapshot hence contains $p = 64^3$ degrees of freedom. Although the time-step size varies between 100 and 10 000 years, this variability does not affect the STP prediction as long as time steps remain consistent across episodes. We use 400 trajectories from the training set (with an initial gas-particle mass of one solar mass) to construct the STP mode basis, and the 50 trajectories from the test set for error analysis. The trajectory visualized in figure 2 at five representative time instances illustrates the evolution of the complex shell structure during the rapid expansion.

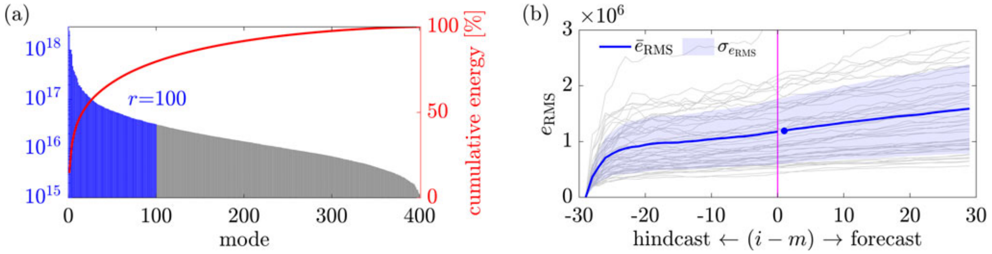

Figure 4. STP prediction of the supernova simulation with a hindcast horizon of $n = 30$ and forecast horizon of $m = 29$. (a) Hindcast mode variance; (b) prediction error, where grey lines represent individual forecasts, the blue line indicates the mean, the solid circle indicates the first time step of the forecast and the light blue band shows the standard deviation. The leading $r = 100$ retained modes, out of a total of 400, capture approximately 80% of the total variance. In (b), the time-step index i is used because the time steps are non-uniform.

The supernova explosion is a transient phenomenon characterized by large-scale structural changes and non-repeating dynamics. Each trajectory captures distinct, non-trivial, anisotropic and evolving shell formations. Our standard forecasting task for such transient data involves predicting the remaining m time instances of the forecast horizon, given an observed hindcast horizon of n time instances. For the first demonstration, we set the hindcast interval to $n = 30$ and predict the remaining $m = 29$ snapshots. Figure 4a shows the hindcast eigenvalue spectrum computed according to equation (2.6). The most correlated structure captured by the leading mode corresponds to a correction to the ensemble mean, as later discussed in the context of figure 6. After a sharp initial drop in mode variance over the first few modes, the eigenvalues exhibit a near-algebraic decay over a large portion of the spectrum. The first $r = 100$ modes capture approximately 80% of the total variance and 344 are needed to reach 95%, as can be seen in figure 4a.

Figure 4b shows the RMS errors for the 50 test trajectories, as well as their mean, defined by equation (2.14), for the forecasted temperature fields. The standard deviation of the error indicates the spread in forecast accuracy across different trajectories. The error is plotted over the time-step index relative to the start of the forecast horizon, which the magenta vertical line separates from the hindcast horizon. Initially, the error increases rapidly over approximately the first 10 time instances of the hindcast, then continues to grow more slowly. Notably, there is no sudden change at the transition from hindcast to forecast, suggesting that the hindcast error is a good indicator of the forecast error.

For a fixed hindcast horizon, the only parameter in this method is the rank r of the STP basis, that is, the number of retained STP modes. Figure 5a shows the mean forecast error, computed from 50 test episodes, for varying r at a fixed $n = 30$. In this dataset, unlike the second example, the smallest error is achieved using the full rank of $r = 400$. However, a substantial reduction to 250 modes only slightly increases the error. By contrast, the rank-1 reconstruction yields the largest error, though it does not differ substantially from higher-rank forecasts, as the ensemble mean already represents the most energetic feature. Figure 5b shows that the forecast error for varying hindcast lengths largely follows the expected trend: longer hindcast horizons lead to more accurate forecasts. The only exception is when $n = 3$, which results in a significantly larger forecast error.

Figure 6 presents the first 10 STP modes, $\boldsymbol{\Phi}^{\star}_{\pm}$, out of a total of 400, at four representative time instances, two from the hindcast horizon and two from the forecast horizon. The ensemble mean, which is subtracted from the data, represents the average rapid expansion of the supernova burst and consequently accounts for the largest fraction of the variance. The first mode similarly exhibits an expanding spherical pattern, providing a correction to the ensemble mean. The remaining modes build a basis that captures the variability of individual trajectories about

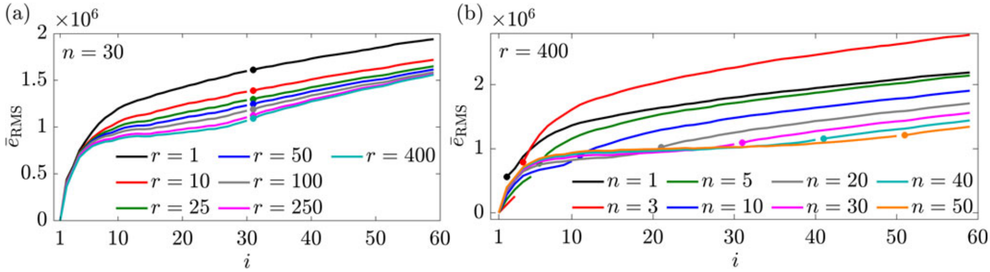

Figure 5. Study of the prediction error in the supernova prediction. (a) Mean prediction error for different numbers r of retained modes, with a fixed hindcast interval length of $n = 30$ and (b) mean prediction error as a function of hindcast interval length n without basis truncation. In all cases, the entire trajectory is considered, such that $n + m = 59$. Solid circles indicate the first time step of the forecast.

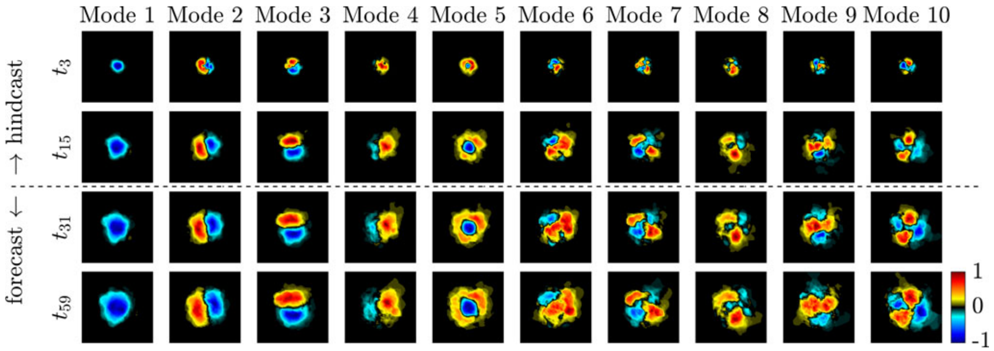

Figure 6. STP modes $\boldsymbol{\Phi}^{\star}_{\pm}$ of the supernova data for a hindcast horizon of $n = 30$ and a forecast horizon of $m = 29$. A prediction using this basis is shown in figure 7. The domain is the same as in figure 2, with $(y, z) \in [-1, 1] \times [-1, 1]$ at $x = 0$. The fluctuating temperature fields are normalized by their maximum absolute values.

the mean, forming a hierarchical structure of multipoles with increasing geometric complexity. Higher modes also tend to be noisier, suggesting reduced convergence.

Figure 7 displays four representative supernova prediction trajectories, obtained using a hindcast horizon of $n = 30$, a forecast horizon of $m = 29$ and a full-rank STP basis with $r = k = 400$. Specifically, we selected trajectories corresponding to the lowest and highest forecast errors at the end of the forecast horizon, as well as two representative episodes that exhibit distinct features. A comparison between the ground truth, i.e. the test data, and the predictions reveals that the primary topological features of the individual explosions are well captured across both the hindcast and forecast horizons. However, certain localized features, such as the nearly perfect circular spots observed in the test data associated with growth from localized nuclei, appear blurred in the predictions. This is expected, given that such localized structures are challenging to represent accurately with a modal basis derived from very limited training data exhibiting substantially different characteristics. Considering these limitations, the overall evolution of these diverse trajectories is, arguably, captured remarkably well.

(b) Open cavity flow

The second example is a time-resolved PIV (TR-PIV) experiment performed by Zhang *et al*. [38] to obtain the two-dimensional velocity field in the centre plane of a Mach 0.6 turbulent flow over an open cavity with length-to-depth and width-to-depth ratios of $L/D = 6$ and $W/D = 3.85$, respectively. We denote the streamwise and wall-normal velocities by $u(x, y, t)$ and $v(x, y, t)$,

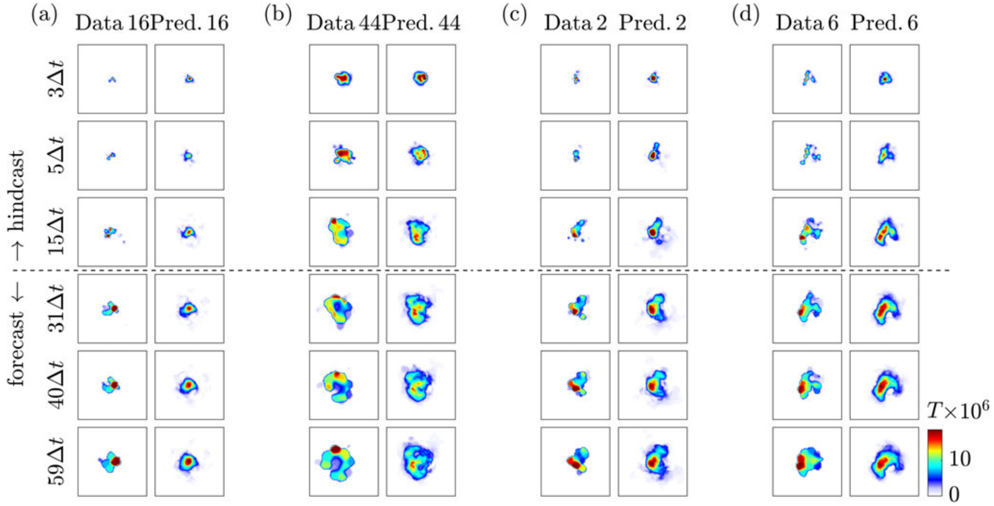

Figure 7. Prediction of the supernova using a hindcast horizon of $n = 30$ and a forecast horizon of $m = 29$. (a) Trajectory with the lowest mean prediction error; (b) trajectory with the highest mean prediction error and (c,d) trajectories with intermediate errors, selected to illustrate variability. The visualization domain is the same mid-cut plane as in figure 2, with $(y, z) \in [-1, 1] \times [-1, 1]$ at $x = 0$.

respectively, and define the state for the ith snapshot as

$$\mathbf{u}_i = \begin{bmatrix} u'_i \\ v'_i \end{bmatrix},$$

where the prime indicates the mean-subtracted (fluctuating) component of each velocity component, flattened and stacked into a single vector. Since the flow field is stationary, the ensemble and temporal means are equivalent, and we can subtract the latter. The database comprises 16 000 flow fields acquired at 16 kHz in a field of view resolved by a 160×57 grid. To eliminate regions with missing data, we removed three points from each lateral boundary, resulting in snapshots with $p = 18\,240$ degrees of freedom each. Further details on the experimental set-up and hardware are reported in [38].

This flow differs fundamentally from the first example because it is statistically stationary rather than transient, and in that it is inherently noisy. Figure 3 illustrates these typical characteristics of a turbulent stationary flow. The instantaneous velocity fields shown in figure 3a,b appear chaotic, and the power spectrum in figure 3c computed at $(x, y) = (5, 0)$ is broadband with several distinct tones emerging from the turbulent background. The two prominent peaks labelled R2 and R3 at frequencies $f_{\mathrm{R2}} = 938$ Hz and $f_{\mathrm{R3}} = 1531$ Hz correspond to the second and third Rossiter tones, stemming from a resonance between downstream-travelling shear-layer instability waves and upstream-travelling acoustic waves [43]. The frequency–wavenumber diagram in figure 3d further enables approximation of the wavenumbers, $k_{x,\mathrm{R2}} \approx 74\,\mathrm{m}^{-1}$ and $k_{x,\mathrm{R3}} = 105\,\mathrm{m}^{-1}$, and thus the phase velocities, $c_{ph,\mathrm{R2}} = 80\,\mathrm{m\,s}^{-1}$ and $c_{ph,\mathrm{R3}} = 92\,\mathrm{m\,s}^{-1}$, of these vortical structures. From these values, we determine that the waves require approximately $32\Delta t$ and $28\Delta t$ to traverse the cavity.

When predicting a stationary flow, where statistics remain constant over time, we have more flexibility in choosing a hindcast horizon. However, the presence of large-scale coherent structures, such as the two dominant Rossiter modes, helps guide this choice and sets expectations for a realistic forecast horizon, even though we later show that short horizons yield the highest prediction accuracy.

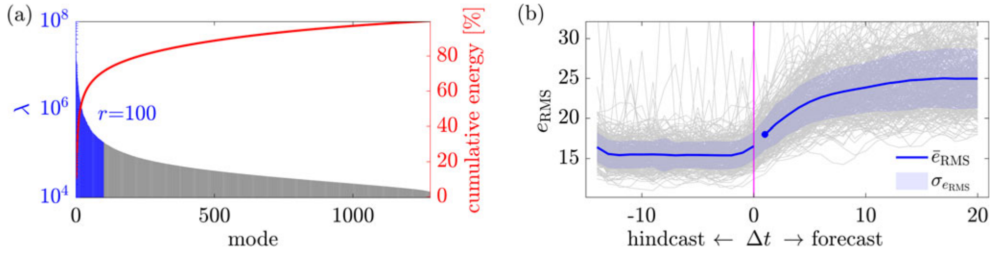

Figure 8. STP prediction of the cavity flow with a hindcast and forecast horizon of $n = 15$ and $m = 20$. (a) Hindcast mode energy and (b) prediction error, where grey lines represent the 316 individual forecasts, the blue line indicates their mean, and the light blue band their standard deviation. The leading $r = 100$ retained modes, out of a total of 1278, capture 71% of the total energy.

Before investigating the dependency of the forecast error on n and r in more detail, we begin with a hindcast horizon of $n = 15$, a forecast horizon of $m = 20$ and retain the leading $r = 100$ modes as a baseline. The data are split in the standard 80:20 manner, and overlapping segments are employed to obtain a sufficiently large number of training and testing episodes, a common practice for stationary data borrowed from spectral estimation [44]. Specifically, each new episode is initiated 10 time steps after the start of the previous episode, resulting in a training ensemble of size $k = 1278$ and 316 testing episodes for independent forecast error quantification. Figure 8a presents the hindcast energy spectrum. Since the state comprises the fluctuating velocity components in the PIV plane, the eigenvalues directly capture the resolved turbulent kinetic energy. The mean prediction error, shown in figure 8b, exhibits distinct behaviour for the hindcast and forecast segments: the hindcast error remains nearly constant, with minor boundary effects, while the forecast error monotonically increases for approximately 10 time steps before settling at a constant level. At this point, the variance of individual trajectories is not accurately captured and the predictive capacity is reached.

Figure 9 examines how the hindcast window length affects forecast accuracy. Figure 9a,b plot the mean RMS error for eight hindcast lengths ($n = 1$–200) as functions of the hindcast window length and the forecast lead time, respectively. Two trends emerge. First, for short lead times, the error increases as n grows: with a fixed number of retained modes, lengthening the hindcast window forces the basis to represent a broader segment of dynamics, which decreases its ability to fit near-term phase and amplitude. Second, the optimal n depends on the desired lead time: $n = 1$ minimizes error for $m \leq 6$, $n = 5$ for $7 \leq m \leq 11$, $n = 10$ for $12 \leq m \leq 15$, and $n = 25$ over most of the remaining horizon. We interpret this dependence in terms of a finite spatio-temporal correlation time: accurate forecasts require a window long enough to capture the correlated wavepackets that drive the evolution, but not so long that largely decorrelated content dominates the reconstruction. An upper bound for the useful window is set by the convective traverse time of the dominant Rossiter modes, approximately 28–32Δt (cf. figure 3). Consistent with this picture, the forecast-error curves in figure 9b plateau once the lead time approaches this scale; for lead times well beyond it, errors grow as spatio-temporal correlations decay. In the limit $n = 1$, where space–time POD reduces to space-only POD, the basis is tuned to the instantaneous field and excels at very short lead times.

We next examine how basis truncation affects forecast accuracy by varying the number of retained modes, $r \leq k$. For fixed hindcast horizons $n = 15$ and 1, and a fixed forecast horizon $m = 20$, the hindcast errors, shown in figure 10a for $n = 15$ and figure 10c for $n = 1$, decrease monotonically with r, with the best reconstruction at full rank. This monotonic improvement follows from the least-squares optimality of the hindcast POD reconstructions. As for variations of the hindcast interval discussed in figure 9, the effect of basis truncation on the forecast error is more nuanced. For forecasting with $n = 15$ in figure 10b, high-rank reconstructions with $r \geq 500$ perform best at very short lead times, whereas intermediate values $25 \leq r \leq 100$ perform better

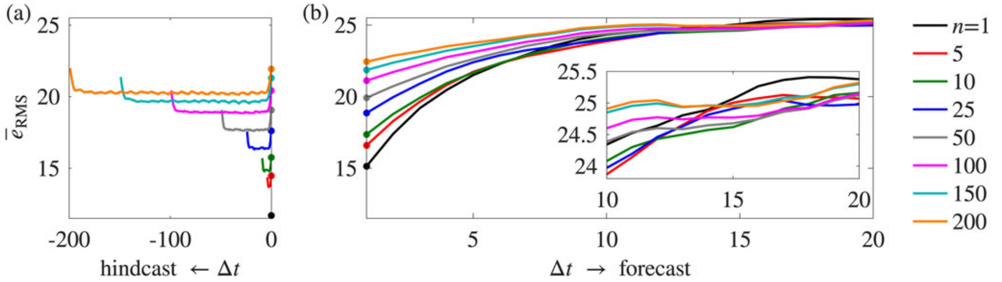

Figure 9. Mean prediction error of the cavity flow as a function of hindcast interval length n. (a) Hindcast error and (b) forecast error. The forecast horizon is fixed at $m = 20$ and the leading $r = 100$ modes are retained.

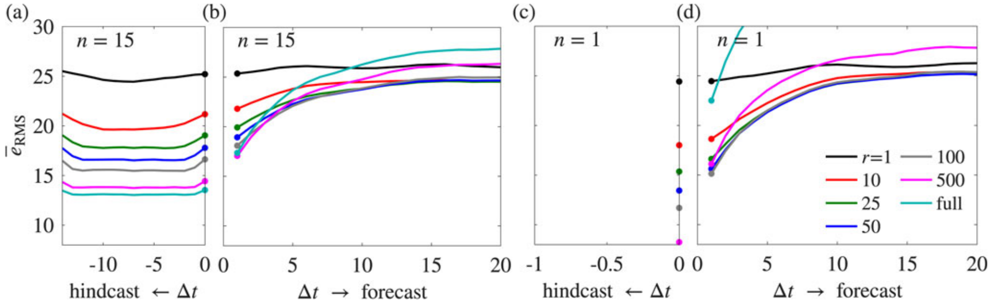

Figure 10. Influence of basis truncation on prediction error for the cavity flow, shown as functions of the number of retained modes r. (a) Hindcast error with $n = 15$; (b) forecast error with $n = 15$; (c) hindcast error with $n = 1$ and (d) forecast error with $n = 1$. The forecast horizon is fixed at $m = 20$.

for lead times $\geq 5\Delta t$. For the $n = 1$ case shown in figure 10d, the same intermediate numbers of retained modes $25 \leq r \leq 100$ perform best for all forecast lead times, with $r = 100$ performing best when less than or equal to $2\Delta t$, albeit by a small margin. In both cases ($n = 15$ and 1), the extremes of rank-1 and full rank are generally suboptimal for forecasting: too few modes underfit and miss essential coupled structures, while very many modes introduce low-energy, poorly conditioned components that inject variance into the learned forecast mapping and overfit transient or noise-dominated content, which degrades medium- and long-lead predictions. At $n = 15$, the brief near-term advantage of very high r arises from reconstruction optimality on the hindcast window, which yields very small errors that carry into the first few forecast steps; beyond that, the many low-energy modes included at high r inflate variance and overfit transient or noise-dominated content, so errors accumulate and intermediate r performs best.

Figure 11 shows the time evolution of the v'-velocity component of four representative STP modes at several time instances spanning both the hindcast and forecast horizons. The first mode represents a slow modulation of the mean field, with its strongest contributions near the back wall and the bottom of the cavity. Its spatial structure suggests a link to centrifugal modes, a well-known low-frequency recirculation phenomenon in open cavity flows [45]. Modes 2 and 5 capture the dominant spatio-temporal structures in the cavity, corresponding to the Rossiter modes highlighted in figure 3c,d. As hydrodynamic instability waves, these structures are highly coherent, energetic and oscillatory, which explains their prominence in the STP modes and their contribution to forecasting. In contrast, mode 50, representative of low-energy, higher-order modes, still displays a discernible structure but appears much noisier. As shown in figure 10, including even higher modes, which tend to be increasingly less converged and noisy, can be detrimental to forecast quality; conversely, truncating these modes can help facilitate noise rejection.

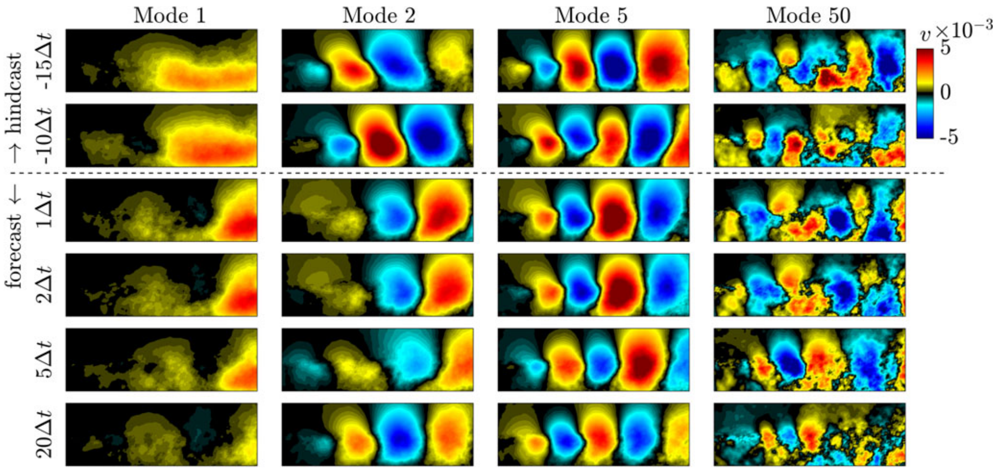

Figure 11. STP modes of the cavity flow for a hindcast horizon of $n = 15$ and a forecast horizon of $m = 20$. A prediction using this basis is shown in figure 12. The domain is the same as in figure 3.

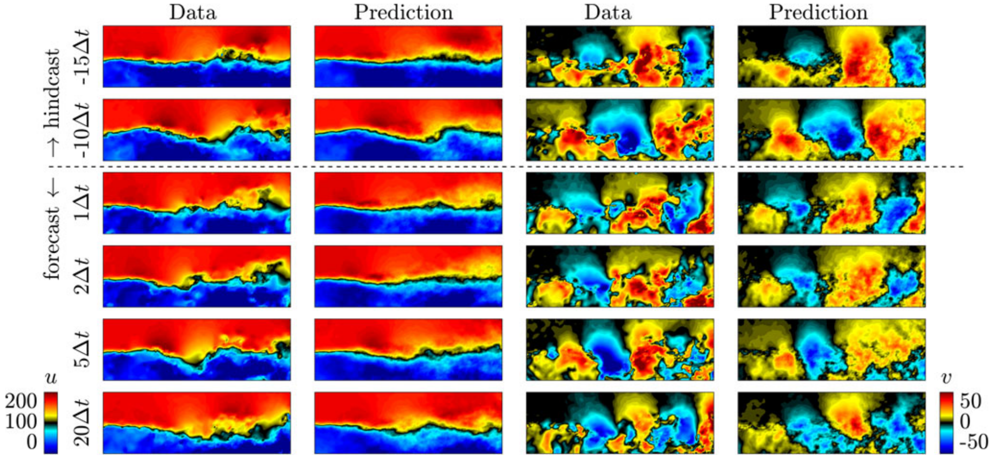

Figure 12. Prediction of the cavity flow using a hindcast horizon of $n = 15$, a forecast horizon of $m = 20$ and $r = 100$ modes retained. The first episode of the test dataset is shown. The domain is the same as in figure 3.

Figure 12 presents the forecast for a single episode, the first one in the test dataset, using the baseline parameters, $n = 15$, $m = 20$ and $r = 100$. The ground truth (i.e. the testing data) is compared to the forecast at the same time instances as in figure 11. To facilitate interpretation, the temporal mean fields have been added back to the velocity fields. At the two representative time steps within the hindcast horizon, the forecast closely resembles the data, at least up to the level of detail the truncated modal basis computed from the training data is able to capture. Over the forecast horizon, the correspondence is less precise; however, particularly for the v-component, which accentuates large-scale coherent patterns, the spatio-temporal evolution of the instability waves convecting along the cavity can still be tracked. Even at the final forecast time step, $20\Delta t$, the dominant flow structure visible in the data, a wave spanning approximately three wavelengths across the field of view, is discernible in the forecast with the correct phase and approximate amplitude. While the turbulent and chaotic nature of the flow inherently limits predictability, the forecast's ability to capture the evolution of these large-scale structures is noteworthy.

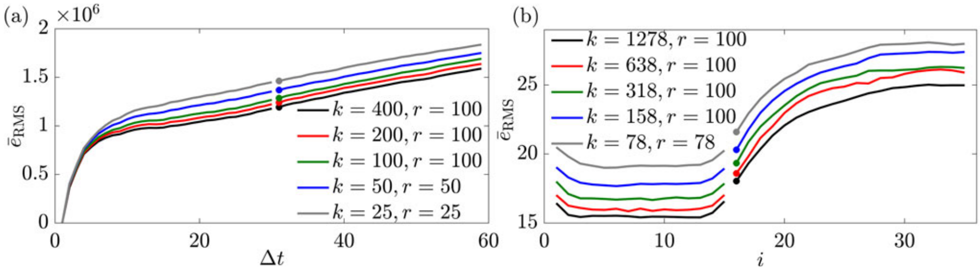

Figure 13. Sensitivity to training ensemble size. (a) Supernova explosion and (b) cavity flow. The hindcast and forecast horizons are fixed, while the training ensemble size is systematically reduced by halving the available training data. The truncation rank is set to $r = 100$; when $k \leq 100$, full-rank prediction with $r = k$ is employed.

(c) Performance

We study two aspects of the method's performance. First, we examine how forecast accuracy depends on the amount of available training data, evaluating this sensitivity for both datasets. Second, we compare the STP forecast error with that of LSTM regression neural networks. This comparison is performed only for the stationary cavity data, where the same set of spatial-only POD modes can be used for both hindcasting and forecasting. In contrast, the transient nature of the supernova data would require a different dimensionality-reduction strategy for a simple LSTM model to be applicable, which is beyond the scope of this work.

In the training ensemble size sensitivity study in figure 13, we use the same baseline hindcast/forecast horizons as in §3a,b (supernova and cavity, respectively). Specifically, for the supernova case, we set $n = 30$ and $m = 29$, whereas for the cavity case, we use $n = 15$ and $m = 20$. The truncation rank is fixed at $r = 100$, except when the ensemble size k is smaller than 100, in which case the prediction is performed at full rank, $r = k$. Similar trends are observed in both datasets: as expected, the reconstruction accuracy decreases with fewer training episodes, yet the relative hindcast accuracy remains a reliable qualitative indicator of forecast accuracy. Notably, important trends, such as the practical forecasting horizon for the cavity flow being limited to approximately 10–15Δt, are consistently observed even with the smallest amount of training data used, that is, 78 episodes for the cavity flow and 25 episodes for the supernova case. For both datasets, each halving of the training ensemble size results in a roughly equal incremental increase in prediction error. The relatively larger errors observed for the smallest ensemble size, particularly pronounced for the supernova data in figure 13a, can be attributed to simultaneous rank reduction necessitated by fewer available training episodes. The consistent increase in prediction error with reduced training data strongly suggests that the ability of the STP basis to represent the data is directly linked to the statistical convergence of the sample covariance matrix. Since the eigenvalues measure the variance (or energy) captured by each mode, constructing a robust and converged basis requires a sufficiently large ensemble to accurately estimate these eigenvalues and their corresponding modes. With fewer training samples, the basis estimation becomes less reliable, systematically degrading forecast accuracy. In summary, this study underscores the importance of using sufficiently large training ensembles for accurate forecasting, but also highlights the robustness of the STP method even when data availability is limited. Next, we compare the STP forecast error to that obtained from DMD and standard LSTM regression neural networks.

DMD, first introduced by Schmid [21], starts from two time-shifted snapshot matrices,

$$\mathbf{U}_1^{n-1} = [\mathbf{u}_1^{[j]}\mathbf{u}_2^{[j]} \cdots \mathbf{u}_{n-1}^{[j]}] \quad \text{and} \quad \mathbf{U}_2^{n} = [\mathbf{u}_2^{[j]}\mathbf{u}_3^{[j]} \cdots \mathbf{u}_n^{[j]}], \tag{3.1}$$

from which we compute the DMD mode matrix $\boldsymbol{\Phi}_{\mathrm{DMD}}$ and the discrete-time eigenvalues $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ (e.g. [46]). Note that the DMD is independently computed for each episode j using the pairs $\{\mathbf{U}_1^{n-1}, \mathbf{U}_2^{n}\}$. DMD forecasts are obtained by advancing the DMD expansion from the initial state at the end of the hindcast window as

$$\mathbf{u}_{n+i} \approx \boldsymbol{\Phi}_{\mathrm{DMD}} \boldsymbol{\Lambda}^{i} \mathbf{b}, \quad i = 1, \ldots, m. \tag{3.2}$$

The vector of complex initial amplitudes is obtained by least-squares projection,

$$\mathbf{b} = \boldsymbol{\Phi}_{\mathrm{DMD}}^{\dagger}\, \mathbf{u}_n^{[j]}, \tag{3.3}$$

with $\mathbf{u}_n^{[j]}$ being the last time step of the hindcast horizon (the initial condition of the forecast) and $(\cdot)^{\dagger}$ the pseudoinverse. Note that the DMD mode basis is computed directly from the hindcast data. In its standard form, DMD requires no separate training. In fact, classical DMD modes and their spectrum do not converge as the sample size increases for stationary data [17]. Although not required for DMD, we retain the same partition in samples indexed by the same superscript $[j]$ for consistency with STP. We found that for DMD, a hindcast horizon of $n = 15$ is near-optimal: smaller values increase forecast error, while larger values offer no further improvement.

For LSTM, we use MATLAB's standard implementation [47], as it is widely accessible, well-documented and reproducible. As direct prediction in physical space is computationally prohibitive, we instead perform predictions on POD coefficients, that is, we use space-only POD for dimensionality reduction. For the stationary cavity flow dataset, the space-only POD modes are computed using the full training dataset, represented by the $p \times K$ snapshot matrix $\mathbf{U}_1^K = [\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \cdots \;\; \mathbf{u}_K]$ where $K = 12\,800$ is the total number of training snapshots, differing from the number of training episodes, $k = 1278$, owing to the overlapping segmentation described in §3b. As before, the state vector includes both velocity components. The LSTM network is trained on the corresponding temporal POD coefficients as latent variables, with rank truncation levels set to $r = 25, 100$ and 500. These levels capture approximately 68%, 83% and 93% of the total variance, respectively, compared to 71% for $r = 100$ hindcast modes. This should, in principle, guarantee that the LSTM network is not disadvantaged by the choice of rank truncation. The LSTM configuration includes a sequence input layer, one LSTM layer with 128 hidden units as default, also varied to 64 and 256 units, and a fully connected output layer. Training targets are generated by shifting the POD coefficients by one time step to forecast future values, and data are normalized to zero mean and unit variance, following standard practices [47]. The network is trained using the Adam optimizer with a constant learning rate of 0.001 for 200 epochs, with data shuffled at each epoch. With the default mini-batch size of 128, the training data are divided into nine mini-batches per epoch, resulting in a total of 1800 iterations.

Figure 14a compares STP with DMD and LSTM in terms of mean prediction error; the corresponding LSTM training losses are shown in figure 14b. Errors are evaluated on full reconstructions of the flow field built from the predicted POD expansion coefficients. We also tested $n = 5$ and 30 and deeper networks with three LSTM layers, with no qualitative change in outcomes. Two trends are evident in the LSTM training loss: (i) the best convergence occurs for the smallest number of modes, $r = 25$, and (ii) for a fixed number of coefficients, increasing the hidden units l generally improves convergence. In figure 14a, the configuration $r = 100$, $l = 128$ yields the smallest short-time error among the LSTM models, while $r = 25$, $l = 128$ performs best for horizons greater or equal to $3\Delta t$. At very short times, performance differences are minimal. Models that perform well at short horizons tend to degrade more at longer horizons, and vice versa. For comparison, we include the STP prediction with $n = 1$ (the configuration giving the smallest short-time error), which achieves lower errors than all tested LSTM architectures for both short and long horizons; only over the intermediate window $3\Delta t \leq t \leq 9\Delta t$ does the LSTM with $r = 25$, $l = 128$ slightly outperform STP, and only by a small margin. We emphasize that our LSTM baseline uses a standard architecture with method-of-snapshots POD for dimensionality reduction to ensure comparable computational cost and reproducibility. While alternative architectures or dimensionality-reduction strategies (beyond the scope here) could yield stronger LSTM results, we similarly did not tune STP beyond selecting the configuration with the smallest short-term error. Classical DMD underperforms STP across

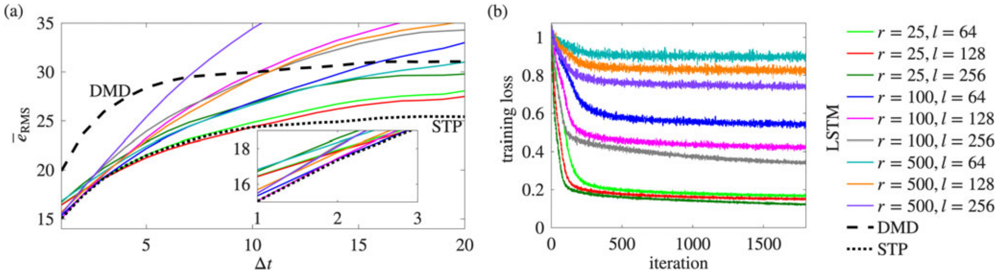

Figure 14. Comparison of STP, DMD and LSTM across varying hidden units, l, and learned POD time coefficients, r. (a) Mean prediction error and (b) LSTM training loss. LSTM and DMD use a hindcast horizon of $n = 15$. For reference, the STP prediction with $r = 100$ and $n = 1$ (the best short-time performance from figure 9) is reproduced. STP and LSTM are trained on the same $k = 1278$ episodes; DMD requires no training.

the entire prediction horizon, surpassing only a few less-well-converged LSTM configurations at longer lead times ($\geq 7\Delta t$). More advanced variants of DMD, in particular ensemble formulations or approaches incorporating time-delayed snapshots (e.g. [48]), may yield better performance.

The primary computational burden of our approach lies in constructing the hindcast correlation matrix and performing its eigenvalue decomposition, or alternatively, the SVD of the hindcast data matrix. Although this can be challenging, particularly in terms of memory usage. For very large datasets, streaming algorithms that are well-established [49,50] can mitigate this burden. To put this into perspective, however, all computations reported in this article were performed using full matrix decompositions of the complete databases (1.1 GB for the cavity flow and 24.9 GB for the supernova), and were completed within minutes for individual computations (excluding parameter sweeps) on a laptop with 64 GB of memory and an Apple M1 Max chip. The compute times of the STP predictions and the LSTM neural networks were comparable.

4. Discussion

We present a method for data-driven forecasting of both transient and stationary phenomena based on an ensemble of realizations of the transient process. Originating from the theory of POD, our method leverages correlation along with the inherent properties of POD, namely, orthogonality and optimality, to generate predictions. For a given ensemble of realizations and a fixed prediction horizon, the only tunable parameters are the rank truncation and the length of the hindcast horizon; no additional hyperparameters are required. The inherent rank reduction and latent space compression are key features of the method, and we contend that the conceptual and algorithmic simplicity of the approach is a major strength.

For both datasets used to test the method, forecasts are produced that appear as continuous extensions of the hindcast horizon into the future. Evaluation on the test data via the conventional 80:20 split revealed that the forecasts effectively capture the evolution of the dominant spatio-temporal patterns in each dataset. The modes, which form a hierarchical basis capturing the variability of the trajectories, exhibit increasing geometric complexity. In the cavity flow case, these modes are clearly linked to physical phenomena that account for their persistence in the data. Furthermore, the hindcast serves as a reliable indicator of forecast accuracy at short prediction times and acts as a lower bound for the forecast error. Notably, only the cavity flow case benefited from rank truncation. This is probably due to the overlapping segmentation used to construct the training episodes and the resulting incorporation of redundant information, which aids the convergence of the leading modes but also inflates the basis with additional, unconverged modes that capture little variance. In contrast, for the supernova data, the full-rank prediction was most accurate. The probable reason is the immense complexity of the expanding shell structures in the individual trajectories and the limited availability of training data to capture the highly anisotropic dynamics characterized by localized structural changes. In this case, the initial expansion stages, emerging from randomly seeded multiple nuclei, determine the subsequent dynamics, and our method successfully predicts the expansion shell pattern over extended forecast horizons.

For a quantitative comparison with state-of-the-art forecasting methods, we evaluated a standard LSTM regression network (predicting POD coefficients) and classical DMD against STP on the cavity-flow case. To ensure fairness, the space–time STP basis and the method-of-snapshots POD basis were chosen to capture comparable energy, and the same training episodes were used. STP achieved the lowest errors at both short and long lead times, while at intermediate horizons (approx. 3–9Δt) the best LSTM configuration was only marginally better. Classical DMD underperformed STP across the entire prediction horizon, surpassing only a few less-well-converged LSTM configurations at longer lead times ($\geq 7\Delta t$). These results highlight the robustness and effectiveness of STP despite minimal tuning; although alternative neural architectures or dimensionality-reduction strategies could narrow the gap, the present evidence supports STP as a strong baseline for future forecasting studies.

While in our examples we predict future states at the same spatial locations and for the same state variables, the approach is not inherently limited to this configuration. The spatial domains of the hindcast and forecast modes can differ, and the variables used for hindcasting need not be identical to those being forecasted. This flexibility opens up applications such as forecasting from sensor data, where the hindcast may differ from the forecast both in terms of the variables of interest and locations. Other promising directions include parametric extensions of the method to account for system variations across different physical or geometric parameters [51], as well as conditional sampling strategies aimed at predicting imminent extreme events based on flow state precursors [15].

Data accessibility. This article has no additional data.

Declaration of AI use. Yes, I have used AI-assisted technologies in creating this article. I used AI-assisted technology (ChatGPT 4o and 5) solely for editorial support: clarifying wording, shortening the abstract and flagging minor inconsistencies in cross-references and figure captions. AI tools were not used to generate scientific content, analyse data, create figures, or write code. All methods, equations, results and conclusions originate from the author. All AI-suggested edits were reviewed and revised by the author, who takes full responsibility for the manuscript's content.

Conflict of interest declaration. I declare I have no competing interests.

Funding. Support from the Air Force Office of Scientific Research (grant no. FA9550-22-1-0541), the Army Research Office (grant no. W911NF-25-1-0222), the National Science Foundation (grant no. CBET-2046311), and the Office of Naval Research (grant no. N00014-23-1-2457) is gratefully acknowledged.

Acknowledgements. Many thanks to Brandon Yeung for suggesting the supernova dataset and providing the MATLAB I/O script, and to Yang Zhang and Lou Cattafesta for generously sharing the cavity flow data.

References

1. Evensen G. 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.: Oceans* **99**, 10 143–10 162. (doi:10.1029/94JC00572)
2. Kalman RE. 1960 A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45. (doi:10.1115/1.3662552)
3. Le Dimet F-X, Talagrand O. 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* **38**, 97–110. (doi:10.3402/tellusa.v38i2.11706)
4. Toth Z, Kalnay E. 1993 Ensemble forecasting at NMC: the generation of perturbations. *Bull. Am. Meteorol. Soc.* **74**, 2317–2330. (doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2)
5. Bauer P, Thorpe A, Brunet G. 2015 The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55. (doi:10.1038/nature14956)
6. Lütkepohl H. 2005 *New introduction to multiple time series analysis*. Berlin, Germany: Springer-Verlag.

7. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. 2015 *Time series analysis: forecasting and control*. Wiley, Hoboken, NJ, USA.
8. Durbin J, Koopman SJ. 2012 *Time series analysis by state space methods*, 2nd edn. Oxford University Press, Oxford, UK.
9. Golyandina N, Nekrutkin V, Zhigljavsky AA. 2001 *Analysis of time series structure: SSA and related techniques*. Chapman & Hall/CRC, Boca Raton, FL, USA.
10. Lumley JL. 1967 The structure of inhomogeneous turbulent flows. In *Atmospheric turbulence and radio propagation* (eds AM Yaglom, VI Tatarski), pp. 166–178. Moscow: Nauka.
11. Lumley JL. 1970 *Stochastic tools in turbulence*. New York: Academic Press.
12. Sirovich L. 1987 Turbulence and the dynamics of coherent structures. I—coherent structures. II—symmetries and transformations. III—dynamics and scaling. *Q. Appl. Math.* **45**, 561–590. (doi:10.1090/qam/910462)
13. Aubry N. 1991 On the hidden beauty of the proper orthogonal decomposition. *Theor. Comput. Fluid Dyn.* **2**, 339–352. (doi:10.1007/BF00271473)
14. Gordeyev S, Thomas F. 2013 A temporal proper decomposition (TPOD) for closed-loop flow control. *Exp. Fluids* **54**, 1477. (doi:10.1007/s00348-013-1477-7)
15. Schmidt OT, Schmid PJ. 2019 A conditional space–time POD formalism for intermittent and rare events: example of acoustic bursts in turbulent jets. *J. Fluid Mech.* **867**, R2–1–R2–12. (doi:10.1017/jfm.2019.200)
16. Frame P, Towne A. 2023 Space-time POD and the Hankel matrix. *PLoS ONE* **18**, e0289637. (doi:10.1371/journal.pone.0289637)
17. Towne A, Schmidt OT, Colonius T. 2018 Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. *J. Fluid Mech.* **847**, 821–867. (doi:10.1017/jfm.2018.283)
18. Schmidt OT, Colonius T. 2020 Guide to spectral proper orthogonal decomposition. *AIAA J.* **58**, 1023–1033. (doi:10.2514/1.J058809)
19. Maurel S, Borée J, Lumley JL. 2001 Extended proper orthogonal decomposition: application to jet/vortex interaction. *Flow, Turbulence and CombustionFlow Turbulence Combustion* **67**, 125–136. (doi:10.1023/A:1014050204350)
20. Boree J. 2003 Extended proper orthogonal decomposition: a tool to analyse correlated events in turbulent flows. *Exp. Fluids* **35**, 188–192. (doi:10.1007/s00348-003-0656-3)
21. Schmid PJ. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28. (doi:10.1017/S0022112010001217)
22. Rowley CW, Mezić I, Bagheri S, Schlatter P, Henningson DS. 2009 Spectral analysis of nonlinear flows. *J. Fluid Mech.* **641**, 115–127. (doi:10.1017/S0022112009992059)
23. Taira K *et al.* 2017 Modal analysis of fluid flows: an overview. *AIAA J.* **55**, 4013–4041. (doi:10.2514/1.J056060)
24. Mezić I. 2013 Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45**, 357–378. (doi:10.1146/annurev-fluid-011212-140652)
25. Chu T, Schmidt OT. 2025 Stochastic reduced-order Koopman model for turbulent flows. (https://arxiv.org/abs/2503.22649)
26. Mandic DP, Chambers JA. 2001 *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. Wiley. Chichester, West Sussex, UK.
27. Hochreiter S, Schmidhuber J. 1997 Long short-term memory. *Neural Comput.* **9**, 1735–1780. (doi:10.1162/neco.1997.9.8.1735)
28. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation. (https://arxiv.org/abs/1406.1078)
29. Lea C, Vidal R, Reiter A, Hager GD. 2016 Temporal convolutional networks: a unified approach to action segmentation. In *Proc. of the European Conf. on Computer Vision (ECCV)*. Springer, Cham, Switzerland.
30. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. 2017 Attention is all you need. In *Advances in neural information processing systems (NeurIPS)*. Curran Associates, Inc., Red Hook, NY, USA.
31. Jaeger H, Haas H. 2004 Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science* **304**, 78–80. (doi:10.1126/science.1091277)
32. Doan NAK, Polifke W, Magri L. 2021 Short-and long-term predictions of chaotic flows and extreme events: a physics-constrained reservoir computing approach. *Proc. R. Soc. A* **477**, 20210135. (doi:10.1098/rspa.2021.0135)

33. Vlachas PR, Byeon W, Wan ZY, Sapsis TP, Koumoutsakos P. 2018 Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proc. R. Soc. A* **474**, 20170844. (doi:10.1098/rspa.2017.0844)
34. Vlachas PR, Pathak J, Hunt BR, Sapsis TP, Girvan M, Ott E, Koumoutsakos P. 2020 Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Netw.* **126**, 191–217. (doi:10.1016/j.neunet.2020.02.016)
35. Raffel M, Willert C, Wereley S, Kompenhans J. 2007 *Particle image velocimetry: a practical guide*, 2nd edn. Springer. Berlin and Heidelberg, Germany.
36. Lario A, Maulik R, Schmidt OT, Rozza G, Mengaldo G. 2022 Neural-network learning of SPOD latent dynamics. *J. Comput. Phys.* **468**, 111475. (doi:10.1016/j.jcp.2022.111475)
37. Hirashima K, Moriwaki K, Fujii MS, Hirai Y, Saitoh TR, Makino J. 2023 3D-spatiotemporal forecasting the expansion of supernova shells using deep learning towards high-resolution galaxy simulations. *Mon. Not. R. Astron. Soc.* **526**, 4054–4066. (doi:10.1093/mnras/stad2864)
38. Zhang Y, Cattafesta LN, Ukeiley L. 2020 Spectral analysis modal methods (SAMMs) using non-time-resolved PIV. *Exp. Fluids* **61**, 1–12. (doi:10.1007/s00348-020-03057-8)
39. Rowley CW, Williams DR. 2006 Dynamics and control of high-Reynolds-number flow over open cavities. *Annu. Rev. Fluid Mech* **38**, 251–276. (doi:10.1146/annurev.fluid.38.050304.092057)
40. Saitoh TR, Daisaka H, Kokubo E, Makino J, Okamoto T, Tomisaka K, Wada K, Yoshida N. 2008 Toward first-principle simulations of galaxy formation: I. How should we choose star-formation criteria in high-resolution simulations of disk galaxies?. *Publ. Astron. Soc. Jpn.* **60**, 667–681. (doi:10.1093/pasj/60.4.667)
41. Iwasawa M, Tanikawa A, Hosono N, Nitadori K, Muranushi T, Makino J. 2016 Implementation and performance of FDPS: a framework for developing parallel particle simulation codes. *Publ. Astron. Soc. Jpn.* **68**, 54. (doi:10.1093/pasj/psw053)
42. Ohana R *et al.* 2024 The well: a large-scale collection of diverse physics simulations for machine learning. In *NeurIPS 2024*. Curran Associates, Inc., Red Hook, NY, USA.
43. Rossiter JE. 1964 Wind tunnel experiments on the flow over rectangular cavities at subsonic and transonic speeds. RAE Technical Report No. 64037.
44. Welch PD. 1967 The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Audio Electroacoust.* **15**, 70–73. (doi:10.1109/TAU.1967.1161901)
45. Brès GA, Colonius T, Crighton DG. 2008 Instability of compressible flow over open cavities. *J. Fluid Mech.* **599**, 309–339. (doi:10.1017/S0022112007009925)
46. Tu JH, Rowley CW, Luchtenburg DM, Brunton SL, Kutz JN. 2014 On dynamic mode decomposition: theory and applications. *J. Comput. Dyn.* **1**, 391–421. (doi:10.3934/jcd.2014.1.391)
47. MathWorks. 2024 Time series forecasting using deep learning. See www.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-learning.html (accessed 13 March 2025).
48. Le Clainche S, Vega JM. 2017 Higher order dynamic mode decomposition. *SIAM J. Appl. Dyn. Syst.* **16**, 882–925. (doi:10.1137/15M1054924)
49. Brand M. 2002 Incremental singular value decomposition of uncertain data with missing values. In *European Conf. on Computer Vision (ECCV)*, pp. 707–720. Springer. Berlin and Heidelberg, Germany.
50. Schmidt OT, Towne A. 2018 An efficient streaming algorithm for spectral proper orthogonal decomposition. *Comput. Phys. Comm.* **237**, 98–109. (doi:10.1016/j.cpc.2018.11.009)
51. Barrault M, Maday Y, Cuong Nguyen N, Patera AT. 2004 An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *C. R. Math.* **339**, 667–672. (doi:10.1016/j.crma.2004.08.006)