

Mapping surface height dynamics to subsurface flow physics in free-surface turbulent flow using a shallow recurrent decoder

Kristoffer S. Moen^{1,*}, Jørgen R. Aarnes^{1,*}, Simen Å. Ellingsen¹ and J. Nathan Kutz^{2,3}

¹*Norwegian University of Science and Technology, Trondheim, Norway*

²*Department of Applied Mathematics, University of Washington, Seattle, Washington, USA*

³*Autodesk Research, 6 Agar Street, London, United Kingdom*



(Received 7 October 2025; accepted 7 July 2026; published 3 August 2026)

Near-surface turbulent flows beneath a free surface are reconstructed from sparse measurements of the surface height variation, by a neural network algorithm known as the SHallow REcurrent Decoder (SHRED). The reconstruction of turbulent flow fields from limited, partial, or indirect measurements remains a grand challenge in science and engineering. The central goal in such applications is to leverage easy-to-measure proxy variables in order to estimate quantities which have not been, and perhaps cannot in practice be, measured. In the application considered here, the aim is to use a sparse number of surface height point measurements of a flow field, or drone video footage of surface features, in order to infer the turbulent flow field beneath the surface. SHRED is a deep learning architecture that learns a delay-coordinate embedding from a few surface height (point) sensors and maps it, via a shallow decoder trained in a compressed basis, to full subsurface fields, enabling fast, robust training from minimal data. We demonstrate the SHRED sensing architecture on two types of turbulent data from recent studies [Aarnes *et al.* *J. Fluid Mech.* **1007**, A38 (2025) and Babiker *et al.* *Phys. Rev. Fluids* **11**, 054802 (2026), respectively]: fully resolved DNS data and PIV laboratory data from a turbulent water tank. SHRED is capable of robustly mapping surface height fluctuations to full-state flow fields up to about one integral length scale deep, with as few as three surface measurements.

DOI: [10.1103/3q12-ylb6](https://doi.org/10.1103/3q12-ylb6)

I. INTRODUCTION

The surface of a gently flowing river is not flat, but comprises features that are characteristic imprints of the turbulent flow beneath it. A qualitative taxonomy groups the coherent, long-lived structures into “dimples,” “boils,” and “scars” [1]. These structures are imprints of surface-attached vortices, upwelling events, and strong surface-tangential vortices [2], respectively. Recent investigations have shown that features of the turbulent free surface elevation which are easily distinguished with computer vision are closely correlated to the turbulent velocity field [3], which is otherwise impractical to measure outside the laboratory. Obtaining estimates of the turbulent velocity field at or close beneath the free surface is of high practical interest. For instance, near-surface turbulence controls the rate at which heat and gas are transferred between water and air [4,5], and the total greenhouse gas discharge from rivers, where these characteristic imprints may be seen, is similar in magnitude to the total gas flux through the ocean surface [6]. *In situ* measurements are slow and expensive, and provide data for a single point or trajectory at a time, whereas observations of the surface, for instance, using airborne drones, can cover larger areas comparatively quickly

*K.S.M. and J.R.A. are to be considered joint first authors.

and at low cost. Such a technique, however, is still at a conceptual stage [7] and will remain so until quantitative predictions of subsurface flow, and concomitant processes such as gas flux, can be made from surface observations only. We present a neural network scheme to directly map surface height variations to the underlying flow fields beneath the surface, thus helping close the gap in practice for using drone sensing.

Data-driven methods have gained traction in turbulence research. Applications include spatial super-resolution or reconstruction from sparse measurements, using shallow networks [8], convolutional neural networks (CNNs) [9], diffusion models [10], and Generative adversarial networks (GANs) [11], as well as physics-informed neural networks (PINNs) as means to reduce nonphysical deviations as accumulated by the preceding networks [12,13]. Furthermore, temporal enhancement of sparse data and reconstruction of turbulence has been attempted using shallow multilayer perceptron (MLP) [13], recurrent neural networks (RNNs) [14], and long short-term memory networks (LSTMs) [15]. In recent years, RNNs, and particularly LSTM networks, have shown strong performance in handling the nonlinear time dynamics of turbulent flows [12,13,15–18]. LSTM networks are a subclass of RNNs, designed to learn long-range dependencies in time series, and improve upon instabilities that may occur in general RNNs due to vanishing gradients in the training optimization. The resulting nonlinear network is powerful in handling sequential datasets, such as time series, and is particularly useful for capturing or forecasting nonlinear temporal dynamics [17]. Traditional linear methods such as Proper Orthogonal Decomposition (POD), Linear Stochastic Estimation (LSE), and linear neural networks are far outperformed by nonlinear networks in most turbulence cases, as pointed out in several papers over the last few decades [12,19–22]. Nonlinear methods are more generalizable across different turbulent cases, whereas traditional linear methods are often limited by restrictive assumptions [23]. LSTM networks have also shown good reconstruction performance with sparse turbulence data [24].

A particular application for which data-driven methods have yielded recent progress is the detailed reconstruction of turbulent flows near interfaces, where measurements at the interface give input to the learning algorithm. Such cases include reconstructing the bottom topography [25] and full subsurface velocity fields [22] from free surface measurements, and reconstructing wall-bounded turbulence from wall measurements only. The latter use CNNs for large-scale reconstruction [26,27], CNN-based variational autoencoders [23], or GANs [20,26], yielding great reconstruction performance compared to linear methods. For the reconstruction of subsurface flow fields from free surface data, Xuan and Shen (2023) [22] applied a CNN on data from DNS, using surface elevation and the surface velocity field as input. The results are promising and outperform previous linear reconstruction models, indicating that turbulent flow fields in, say, a river or lake may indeed be inferred from the water surface only. In this case, however, the CNN is applied image by image for reconstruction, and hence is independent of temporal dynamics. Moreover, it is pointed out that CNN methods may underestimate the fluctuating amplitudes of large-scale structures. Beyond these issues, CNN-based approaches often prove to be computationally expensive and either require dense surface fields or struggle with noisy experimental data [12], leaving a gap for a lightweight model that can operate on sparse field data and generalize across simulation and laboratory flows. A more general issue with “sensing” turbulent flow fields from interface measurements is the fall in accuracy with increased distance from the measured interface, due to less correlation of the flow and interface far from the latter. Hence, closing the gaps of both efficiency and accuracy is essential to take subsurface flow sensing to a level of practical utility, i.e., for practical remote sensing of river turbulence and subsequent gas-exchange estimation.

A recently developed method by Williams *et al.* [28], the SHallow REcurrent Decoder (SHRED), takes inspiration from recent data-driven networks to reconstruct spatiotemporal fields from sparse sensor measurements of a single quantity. It is based on the principle of separation of variables, where the temporal dynamics is learned separately with an LSTM encoder, while spatial structures are recovered with a shallow decoder network (SDN); see illustration in Fig. 1. More specifically, time series of sparse sensor measurements of a single quantity field (e.g., pressure, surface elevation, velocity) are given as input to the LSTM. It then constructs a latent space representation of the

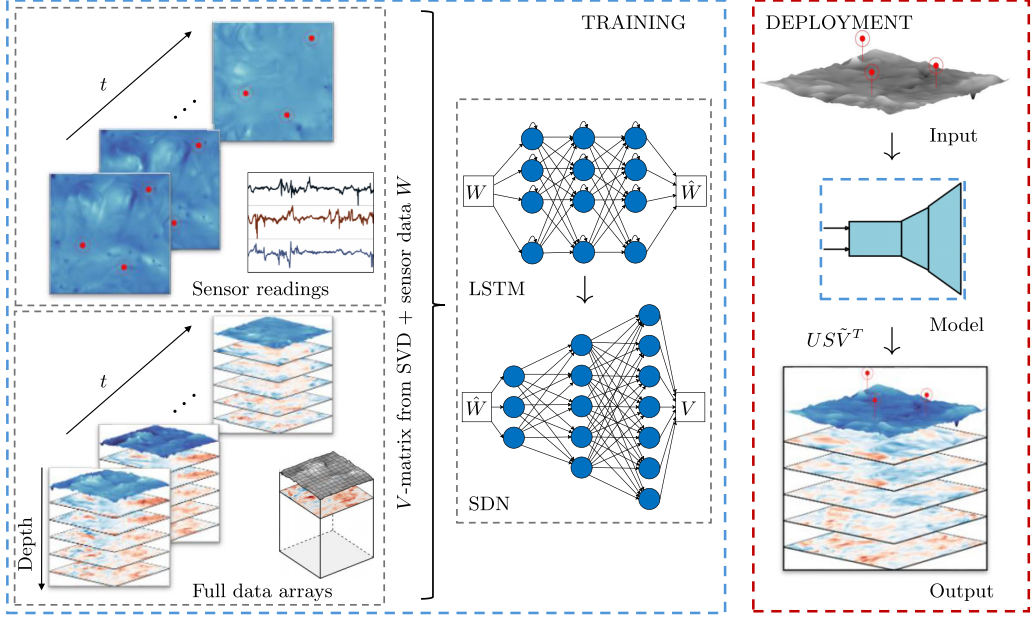


FIG. 1. Illustration of the SHRED architecture, training, and deployment phase. Left: Input and output to SHRED, consisting of time series from three sensors (W) combined with the time dynamics of the DNS or PIV/profilometry data (compressed V -matrix from SVD of slice planes), respectively. Middle: W processed by long-short time memory; the resulting latent representation $\hat{W}$ mapped onto V by the shallow decoder network. Right: The model is deployed by taking new measurements of the surface, outputting the reconstructed time dynamics $\hat{V}$. After processing the full surface and velocity slice planes are reconstructed by matrix multiplication of $\hat{V}^T$ with the compressed SVD matrices U and S .

time dynamics of the field that the SDN is trained to map onto all correlated fields of interest or their compressed representations [29,30]. This is possible because of Takens's embedding theorem, which states that as long as each time series is treated as a delay coordinate embedding, the underlying flow dynamics attractor is contained in the embedding [31]. Hence, this enables the decoder to learn a smooth mapping to the full spatial field. The compressive training is critically enabling as it allows for rapid and robust training on laptop-level computing for a variety of high-dimensional, multiphysics systems such as plasma Hall thrusters [29], nuclear reactors [32], circulating fuel reactors [33], and reduced order models [30]. For the first two applications, the models featured 14 and 21 coupled PDEs, respectively, with only a single field measured with three randomly placed sensors. SHRED can also be used with mobile sensors [34] and can even be combined with the *sparse identification of nonlinear dynamics* (SINDy) or Koopman methods for model identification from sensing alone [35]. The ability to infer and sense fields distinct from the one being measured makes SHRED a suitable candidate in the quest for remote sensing of subsurface flows, as in rivers and oceans. Hence, SHRED is a promising addition to the landscape of data-driven methods for fluid flow sensing. It is therefore of great interest to apply SHRED in the context of free-surface turbulence and subsurface reconstruction.

This paper presents a case study of SHRED deployed to flow data from turbulent free-surface flow, made available to us from two separate sources: direct numerical simulations performed at the University of Minnesota, previously published in [2,36], and experiments from a recent laboratory study at the Norwegian University of Science and Technology [37,38]. In both datasets, isotropic, homogeneous turbulence was created well below the surface whence it naturally diffuses towards the surface. Figure 1 exhibits the SHRED algorithm in relation to our mapping of

surface measurements to subsurface flows. This category of free-surface turbulent flow is well documented in the turbulence literature, both experimentally [39–47] and numerically [2,3,48–50], and can be considered fundamental to our understanding of free-surface turbulence dynamics. The relevance to a number of naturally occurring flows is immediate, such as the near-surface flow of a river or the upper layer of the ocean when the free-surface shear stress due to wind is negligible. We present the performance of SHRED on such free-surface turbulent flow data as a proof of concept of data-driven sensing of subsurface flows from surface measurements only, in a highly challenging turbulent environment characterized by large intermittent structures and absence of periodic features. In stress testing SHRED for different Reynolds numbers from DNS and laboratory experiments, we aim at connecting the SHRED performance to real nonideal flows as typically found in natural environments such as rivers and oceans. Moreover, our results highlight the significant experimental gain of reconstruction of subsurface flows from surface-only measurements, as a step towards more accurate nonintrusive measurement techniques.

The paper is organized as follows: In Sec. II we describe the reconstruction model SHRED, and in Sec. III we summarize the simulations and experiments from which our training data and reconstruction datasets originate, as well as data compression and performance metrics used. We present our results in Sec. IV, among them a set of error metrics which discloses how the performance of SHRED varies with distance from the reconstructed plane to the free surface. Lastly, we draw conclusions in Sec. V and include an Appendix with details on parameter tuning thereafter.

II. SHRED: SHALLOW RECURRENT DECODER

SHRED [28] is a generalization of the separation of variables methods for solving partial differential equations (PDEs) [51]. Separation of variables is also the underlying technique used for many spectral integration methods for solving PDEs [52]. The method assumes that a solution can be separated into a product of time and space functions $u(x, t) = T(t)X(x)$. The solution reduces the PDE into an ordinary differential equation for time $T(t)$ and a boundary value problem for space $X(x)$.

To demonstrate the method, consider the constant coefficient linear PDE

$$\dot{u} = \mathcal{L}(\partial_x, \partial_x^2, \dots)u, \quad (1)$$

where $u(x, t)$ specifies the spatiotemporal field of interest subject to the physics imposed by the operator $\mathcal{L}$. Sample initial conditions (IC) and boundary conditions (BCs) are given by

$$\text{IC: } u(x, 0) = u_0(x), \quad (2a)$$

$$\text{BCs: } \alpha_1 u(0, t) + \beta_1 \partial_x u(0, t) = g_1(t) \quad \text{and} \quad \alpha_2 u(L, t) + \beta_2 \partial_x u(L, t) = g_2(t). \quad (2b)$$

This may be generalized to systems of several spatial variables, a system with no time dependence, or a coupled system of equations. The linear operator $\mathcal{L}$ specifies the spatial derivatives, which in turn model the underlying physics of the system. Simple examples of $\mathcal{L}$ include $\mathcal{L} = c\partial_x$ (the one-way wave equation) and $\mathcal{L} = \kappa\partial_x^2$ (the heat equation) [52].

The earliest solutions of linear PDEs assumed separation of variables whereby $u(x, t) = \exp(\lambda t)X(x)$ was a product of a temporal (exponential) function multiplied by a spatial function. The parameter λ is in general complex and specifies the eigenfunction solution

$$u(x, t) = \sum_{n=1}^N a_n \exp(\lambda_n t) \phi_n(x), \quad (3)$$

where $\phi_n(x)$ are the eigenfunctions of the linear operator and λ_n are its eigenvalues [$\mathcal{L}\phi_n(x) = \lambda_n\phi_n(x)$]. Here a finite-dimensional approximation N is assumed, which is standard in practice for numerical evaluation. The solution of Eq. (3) is a general solution which models all possible

solutions. To specify a unique solution, initial conditions $u(x, 0) = u_0(x)$ are typically imposed in order to uniquely determine the coefficients a_n . Specifically, at time $t = 0$, Eq. (3) becomes

$$u_0(x) = \sum_{n=1}^N a_n \phi_n(x). \quad (4)$$

Taking the inner product of both sides with respect to $\phi_m(x)$ and making use of orthogonality gives

$$a_n = \langle u_0(x), \phi_n(x) \rangle. \quad (5)$$

As an alternative to specifying the initial data at all spatial points, SHRED instead specifies measurements at a single spatial (sensor) location x_s , but with a temporal history. Multiple point measurements can be used as well without loss of generality. Thus, if SHRED has N temporal trajectory points, this gives at each time point of the measurement a constraint:

$$u(x_s, t_j) = \sum_{n=1}^N a_n \exp(\lambda_n t_j) \phi_n(x_s) \quad \text{for } j = 1, 2, \dots, N. \quad (6)$$

This results in N equations for the N unknowns a_n . Specifically, the $N \times N$ system of equations $\mathbf{Ax} = \mathbf{b}$ is prescribed by the vector components $x_k = a_k$ and $b_k = u(x_s, t_k)$ and matrix components $(a_{kj}) = \exp(\lambda_k t_j) \phi_k(x_s)$. As with the initial condition (2a), the time trajectory of measurements at a single location uniquely prescribes the solution. This analysis can easily be generalized to include multiple sensor measurements at a single time point. Thus, if there are two measurements at a given time t_j , then only $N/2$ trajectory points are needed to uniquely determine the solution. Likewise, three sensor measurements at a given time require $N/3$ trajectory points. In addition to stationary sensors measurements, one can also consider mobile sensors whereby the measurement of the system is at different locations over time [34]: $x_s = x_{s(t_j)}$. The above arguments are easily modified so that the vector component $b_k = u(x_{s(t_j)}, t_k)$ and matrix components $(a_{kj}) = \exp(\lambda_k t_j) \phi_k(x_{s(t_j)})$.

Thus temporal trajectory information at a single spatial location, or with a moving sensor, is equivalent to knowing the entire initial condition. SHRED provides a generalization to separation of variables $u(x, t) = T(t)X(x)$ by encoding time with a time sequence model such as an LSTM model and a decoder model for full-state reconstruction of space. Rigorous theoretical bounds of SHRED are difficult to achieve, much like analytic and numerical solutions are difficult to rigorously bound in computational PDE settings. But in the linear limit, the above arguments show explicitly why SHRED is guaranteed to work and recover the full spatiotemporal field exactly.

A. Nonlinear PDEs

For nonlinear PDEs of the form

$$\dot{u} = N(u, \partial_x u, \partial_x^2 u, \dots), \quad (7)$$

numerical methods are commonly used to generate solutions subject to the initial and boundary conditions (2). Consider a spectral solution technique [52] whereby numerical solutions are approximated by a spectral basis

$$u(x, t) = \sum_{n=1}^N a_n(t) \phi_n(x). \quad (8)$$

Typical examples of spectral techniques include using Fourier modes or Chebychev polynomial for $\phi_n(x)$. This spectral decomposition turns the PDE into a systems of N coupled ordinary differential equations for $a_n(t)$:

$$\frac{da_n}{dt} = f_n(a_1, a_2, \dots, a_N) \quad \text{for } n = 1, 2, \dots, N. \quad (9)$$

The solution of the N -dimensional differential equation has N unknown constants of integration that are typically uniquely determined by applying initial conditions and orthogonality in Eq. (8):

$$a_n(0) = \langle u_0(x), \phi_n(x) \rangle. \quad (10)$$

As with the separation of variables solution, we can instead assume that we can construct a general solution for Eq. (9) which has N constants of integration. The constants of integration can be determined by requiring the solution to satisfy N temporal trajectory points, giving at each time point of the measurement

$$u(x_s, t_j) = \sum_{n=1}^N a_n(t_j) \phi_n(x_s) \quad \text{for } j = 1, 2, \dots, N. \quad (11)$$

This gives N constraints for the N unknown constants of integration, thus uniquely determining the evolution of the a_n in Eq. (9). Mobile sensors can also be used to enforce the constraints required for a unique solution. For details on how this extends to coupled PDEs, see [28].

B. Training and deployment

We train and deploy SHRED on the four datasets detailed in Sec. III, each consisting of simultaneous velocities and free-surface elevation resolved in space and time, according to the principles laid out in Fig. 1. We input time series of surface elevation from randomly placed surface sensor points into a two-layer LSTM. The input matrix $\mathbf{W}$ is an $n_s \times n_t$ matrix, where n_s is the number of sensors and n_t is the length of the time series. For all results we report, $n_s = 3$, excepting the discussion of influence of the number of sensors on reconstruction accuracy in Appendix B. The time series length n_t depends on dataset and whether we are in the training or deployment stage. The LSTM encodes these input sequences from the surface sensors into a latent representation of their temporal dynamics, denoted $\hat{\mathbf{W}}$ in Fig. 1, of size $n_{\text{lat}} \times n_t$, with $n_{\text{lat}} = 64$ latent state dimensions. This latent vector is then passed to a shallow decoder network (SDN), which maps it onto the compressed $\mathbf{V}$ -matrices, obtained from SVD and rank truncation, for the surface elevation and the subsurface velocity fields across depth. These fields are used in training and validation not to learn the subsurface time dynamics but only to learn the mapping of the surface time dynamics onto the subsurface fields. This is an essential detail in the context of remote sensing. To output the full spatiotemporal fields, we save the compressed $\mathbf{U}$ - and $\mathbf{S}$ -matrices from the SVD of the training data, and matrix multiplication with the reconstructed time dynamics, $\hat{\mathbf{V}}$, then yields the reconstruction fields. This is computationally cheaper than feeding in the full spatiotemporal fields, although that is also possible.

Each SHRED model is independently trained on one continuous dataset at a time, that is, no cross-case or cross-domain training is used. For each dataset, we randomly split the data into 80% training, 10% validation, and 10% testing snapshots, all within the same time series. An ensemble of SHRED models is trained in each case to assess convergence and uncertainty for each run. Note that using SHRED as a forecasting tool, where the training and test data are separated in a temporal sequential manner, is beyond the scope of the present study. For an example of SHRED used for forecasting, see [53].

SHRED uses a two-layer LSTM encoder and a two-layer SDN decoder (with no dropout). We use the Adam optimizer with an initial learning rate of 10^{-3} , and train on mini-batches of 64 time snapshots. The loss function is the mean squared error (MSE), computed between reconstructed and true $\mathbf{V}$ -matrices in the compressed domain. Details can be found at [54].

Figure 2 shows a typical validation MSE loss curve for datasets S2 and E2 (flow metrics in Table I). We observe a steep drop in validation after around 40 epochs for dataset S2, and around 200 epochs for E2, and a convergence typically occurs after a few hundred epochs. Running on GPU, a full SHRED run (excepting the data compression step) of surface paired with a single

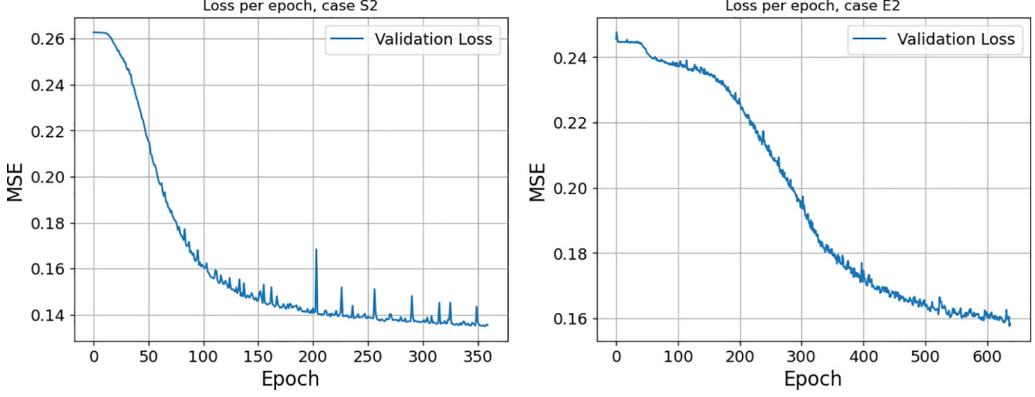


FIG. 2. Example of typical MSE loss profile for the validation datasets S2 (left) and E2 (right).

velocity plane takes 1–2 minutes on a regular desktop computer. A simultaneous run with all planes (only possible with the datasets S1 and S2) takes up to 10 minutes.

III. DATA GENERATION, COLLECTION, AND EVALUATION

A. DNS and experiments

We deploy SHRED to four different datasets, two sets of data from direct numerical simulations and two from experimental data which which utilize a combination of PIV and profilometry to capture both the surface displacement and the subsurface dynamics. The data have been thoroughly documented elsewhere, and only a brief outline is repeated here. Details on the simulation parameters, numerical schemes, and grid resolution criteria based on grid independence studies can be found in [48,55–57], with specifics related to the dataset in use given in [2]. Details of the data capture method in the experiment and the experimental setup can be found in [38,58], respectively.

The simulations mimic experiments of free-surface homogeneous turbulence where turbulence is generated far beneath the water surface and diffuses towards the surface through self-interaction of turbulence vortices and viscous diffusion, with turbulence generated by an oscillating grid (see, e.g., [39,41,44,46]) or jets with zero net flow (e.g. [38,43,47,59]). In the simulations [2,48,56], the nondimensional, incompressible Navier-Stokes equations and continuity equation were solved on a 3D domain, with periodic boundary conditions in horizontal directions, a free-slip boundary at the bottom, and a free surface boundary on the top. The latter is enforced by the dynamic and kinematic boundary conditions and tracked by a surface-adhering grid that undulates in the vertical direction as the surface elevation varies. Turbulence is generated in the center region of the domain by random linear forcing [60,61], modulated by a depth-dependent function which is strongest in

TABLE I. Flow and turbulence properties for data from simulations (S) and experiments (E). From left: case label, turbulent Reynolds number, turbulent Froude number ($Fr_\infty = u' / \sqrt{2gL_\infty}$), turbulent Weber number ($We_\infty = \rho u'^2 L_\infty / \sigma$), Taylor Reynolds number, integral length scale, Taylor length scale, and Kolmogorov length scale. Details on the calculation of the tabulated values can be found in [38].

Dataset	Re_∞	Fr_∞	We_∞	Re_λ	L_∞ [mm]	λ_T [mm]	L_K [mm]	u_{RMS} [mm/s]
S1	267	0.009	∞	47	21.7	7.59	0.56	6.15
S2	782	0.010	∞	84	43.0	9.18	0.51	9.10
E1	7636	0.040	5.2	279	77.1	5.5	0.17	49.5
E2	17 369	0.050	18.1	397	115.4	5.3	0.13	75.3

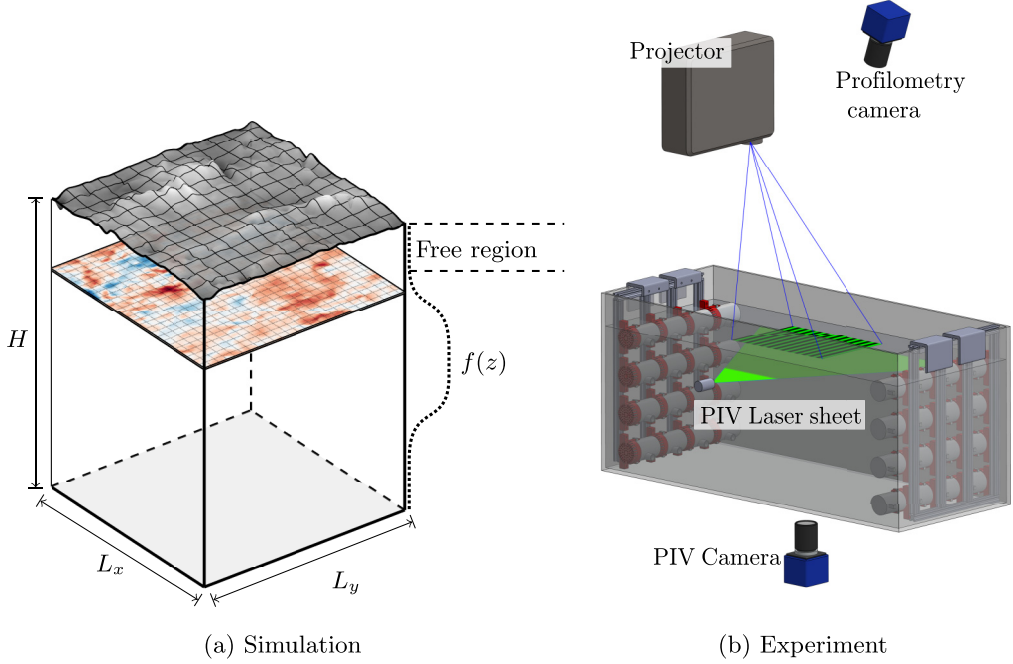


FIG. 3. (a) Schematics for the simulation domain, with undulating grid following the free surface, depth-dependent forcing function $f(z)$ ($= 0$ in the free region), and periodic boundaries along the horizontal directions. Adapted from Fig. 2 of [2]. (b) Facility used in experiments, reprinted from Ref. [38] with permission. Turbulence was generated by randomly actuated jets, and data are captured by profilometry (projector and camera 1) and PIV (sheet-making optics and camera 2; laser and mirrors not in the frame).

the center of the domain and drops to zero well below the surface-influenced region [see Fig. 3(a)]. The top region, denoted the “free region,” is free from random forcing. Only data from the free region are used in the analysis. Surface tension is neglected in the simulations, which has little effect on the degree of correlation between surface and bulk as long as the flow is in the high-Weber-number regime (see extensive discussion in [2]). The mesh sizes of the two DNS datasets are $128 \times 128 \times 348$ and $256 \times 256 \times 660$, for the low and moderate Reynolds number case, denoted S1 and S2 in Table I, respectively.

The experimental flow data are from the experiment of Babiker *et al.* [38] where full details can be found. Turbulence was generated by randomly actuated water jets below the water surface in a rectangular tank [see Fig. 3(b); note that only the two bottom rows of the jets were active]. With a data collection technique developed recently by Semati *et al.* [58] that combines surface profilometry with particle image velocimetry (PIV), the surface displacement and velocity in a subsurface horizontal plane were captured simultaneously. Unlike in the simulations, only a single velocity plane was resolved at a time, serially measuring at four different depth levels between ensembles to get estimates for depth dependency. During characterization of the experiment, a vertical laser sheet was used, and, hence, well-resolved root-mean-square measures of the velocity are obtained. The experimental dataset consists of the measured free-surface elevation directly above each measured PIV velocity field for 20 time intervals, or ensemble cases, each with a 1-minute duration. Of these, 14 were considered suitable data for the SHRED analysis we perform here.

Flow metrics for the four datasets are listed in Table I. There two additional dimensionless numbers are included alongside the Reynolds number. These are the Froude number and the Weber number. While the Reynolds number expresses the ratio of inertial to viscous forces, the Froude

number takes into account the gravitational acceleration, g , and the Weber number expresses the influence of inertial forces to forces due to surface tension, σ . Details on how these quantities are calculated in practice can be found in [2]. Extensive details on the comparison of these DNS and experimental datasets by dimensionalization of the DNS data and appropriate estimation of integral scales and reference depths can be found in [38].

For brevity, we limit our scope to reconstructing one component of the velocity field, a horizontal velocity component denoted u . Since the turbulence is horizontally isotropic, the choice of horizontal axis is without consequence. In our analysis, we refer to the root-mean-square velocity as a representative quantity, which we define as

$$u_{\text{RMS}} = \left(\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n u_{x_i, y_j}^2 \right)^{\frac{1}{2}}, \quad (12)$$

where the indices x_i and y_j denote the positions in the horizontal plane, referring to the discrete data matrices in the DNS and PIV data grids, respectively, of dimension $m \times n$. Thus, u_{RMS} varies as a function of depth, z , and time, t . Since we always consider only the single velocity component u , u_{RMS}^2 should not be confused with twice the kinetic energy density.

B. Compression

When reconstructing subsurface turbulent flows, data compression may facilitate good reconstruction and is also justified by the nature of turbulence. We use singular value decomposition (SVD) to evaluate the relevant modes and scales for the compression of the high-dimensional data to a low-rank representation capturing the essential dynamics. The SVD is performed on each plane and flow variable separately, ensuring no data leakage across the depth of the flow from the decomposition.

The DNS and experimental datasets contain turbulence data with a wide range of spatial scales. From turbulence theory, it is well known that energy is injected at the largest scales, before cascading to smaller and smaller scales in the inertial $k^{-5/3}$ range, and dissipating to heat at the Kolmogorov length scale. The majority of the turbulent kinetic energy is carried by the largest structures, which dominate the transport of heat, momentum, and mass in the flow. When we use SHRED to reconstruct the turbulent free-surface flow from surface measurements, we aim to accurately capture the large scales and avoid overfitting to the intermittent and unpredictable small scales of the turbulent spectrum. We achieve this by compressing the data by SVD, keeping only a small amount of the decomposed data. In addition to increased accuracy, using SHRED with rank-reduced data has the advantage of a very significant speed-up of training, validation, and reconstruction. (For details on the SVD algorithm, see, e.g., [17,62].)

Figures 4 and 5 show the SVD modes and turbulence power spectra for datasets S2 (DNS) and E2 (experiment), respectively. Through SVD the flow data are decomposed in matrices that contain spatial ($\mathbf{U}$), temporal ($\mathbf{V}$), and energy ($\mathbf{S}$) information (strictly speaking, the latter matrix contains the singular values, σ_i , which can be considered as a measure of energy for the flow patterns in $\mathbf{U}$ and $\mathbf{V}$). Decomposing a dataset of n time steps results in n modes. For example, dataset S2 has 12 500 modes; E2 has 900. The modes are ordered by energy or singular value, hence $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$. After decomposition, the exact flow dataset can be reconstructed by matrix multiplication of $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}^T$. Spatial and temporal coefficients of eight modes for a velocity field near the surface are shown in rows 1–2 (spatial) and 3–4 (temporal), in Figs. 4 and 5. We observe that the large-scale spatial modes have large singular values, meaning the majority of the energy content is found in these modes—as expected from our knowledge of the turbulent flow. Likewise, modes of higher order (smaller singular value or lower energy) are related to smaller structures, as seen from comparing, e.g., u_2, σ_2 to u_{50}, σ_{50} in Fig. 4. Moreover, we observe a more rapid variation in the temporal coefficients as the order is increased. This is as expected for small compared to large scales in the turbulence,

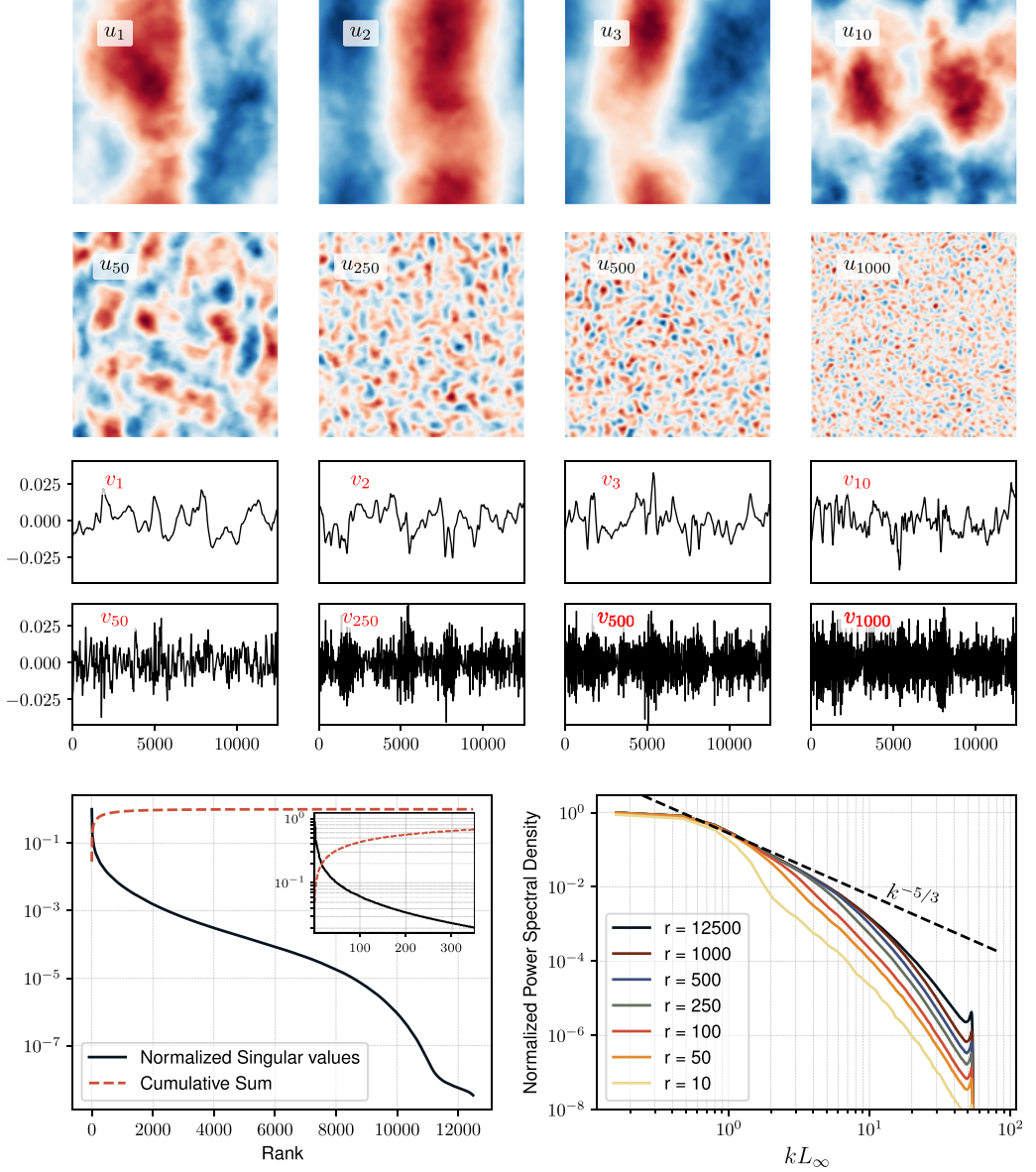


FIG. 4. Top two rows: Eight representative spatial SVD modes (u_i) and temporal modes (v_i) of the decomposed horizontal velocity component u at the surface for the S2 data case, chosen for illustrative purposes, for a snapshot at an arbitrary point in time. Middle two rows: Evolution of the same modes in time (represented by frame number). Lower left: singular values of the SVD modes, with inset showing the first 350 values. Lower right: Normalized turbulent power-density spectrum (PSD) for velocity fields compressed by retaining only modes from 1 to rank r .

yet it may also be a result of noise, which typically shows up in small-scale data, another argument for using a compressed rather than a full dataset with SHRED.

The choice of rank (i.e., the range of modes to retain) for the compressed, low-rank representation of the data is done on the basis of (1) the observance of the spatial and temporal modes and their scales and variance, (2) the rapid decline of singular values per rank number, (3) from the turbulence

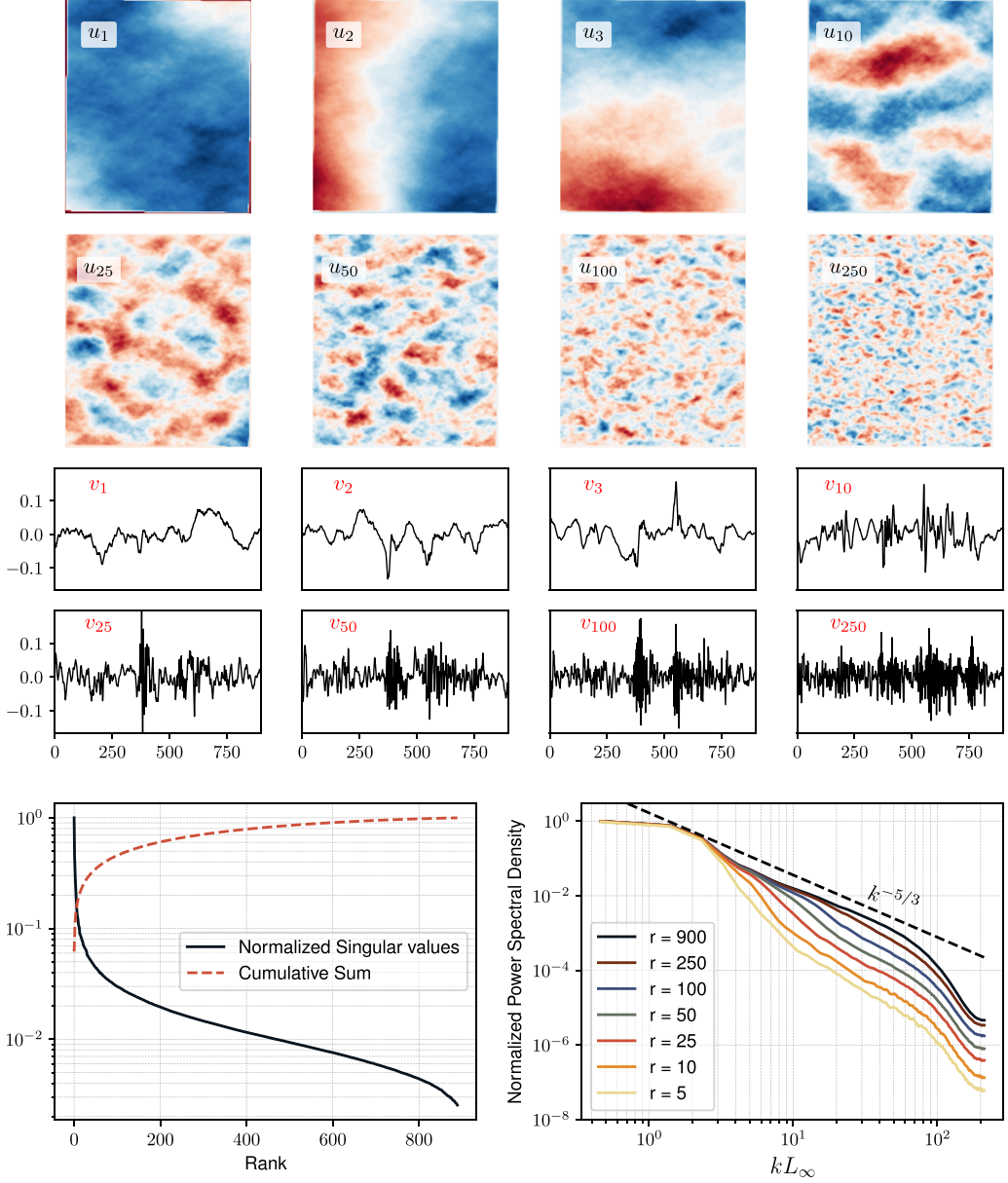


FIG. 5. Same as Fig. 4, but for the high-Reynolds-number experimental dataset E2. SVD modes are found from a horizontal velocity component measured in a plane located 1 cm ($0.14L_\infty$) below the free surface. Note that the total number of SVD modes is 900 here.

spectra dependency on the rank number, and (4) postresult evaluation of optimal values for best SHRED performance. From a computational perspective (points 1 and 2), the optimal low-rank representation contains the relevant modes and covers most of the cumulative sum of the singular values. From a fluid mechanics perspective (point 3), we want to choose a rank of truncation that yields a compressed dataset which matches the original spectrum well into the upper part of the inertial range. The results in Figs. 4 and 5 suggest that rank 250 is sufficient for the S2 dataset, and rank 100 is sufficient for E2. This corresponds to a cutoff at normalized wave number $k_c L_\infty$

TABLE II. SVD compression details. From left: Dataset, number of SVD modes in the uncompressed data, number of SVD modes in the compressed data, normalized cumulative sum of singular values of the compressed data, percentage of modes left out of the compressed dataset, and the upper cutoff wave number where low- and full-rank turbulence spectra start to deviate by more than 10%, normalized by integral length scale.

Dataset	Full-rank	Low-rank	Cumulative sum	Rank truncation	$k_c L_\infty$
S1	10 900	225	75.0%	97.9%	3.6
S2	12 500	250	72.0%	98.0%	3.7
E1	900	100	52.0%	89.9%	6.9
E2	900	100	45.0%	89.9%	10.3

roughly equal to 3–10, where k_c is the wave number where the 1D turbulence spectra of the compressed and uncompressed data first deviate by more than 10% (see Sec. III C 3). Hence, all scales down to a third of the integral length scale for DNS, and sixth to tenth for experiments, are resolved. A parametric study of SHRED performance for different levels of compression is included in Appendix A. We find that the performance of SHRED decreases when the rank we set for compression is higher than a certain value, i.e., when the level of compression is too low, with the exact value being different for each flow. Taking this into account, we choose the low-rank representation of the data to be truncated as listed in Table II. We note that for experimental data cases, we choose a rather low rank, with only a cumulative sum of singular values of 45%–50%. This is because the noise present in the experimental data makes it significantly harder to handle higher-rank data than in the data from the DNS.

C. Performance metrics

When evaluating SHRED performance, we choose a set of depth-dependent error metrics that capture different aspects of the reconstruction error. When reporting results in terms of these metrics in Sec. IV, we set “ground truth” to be the uncompressed, full-rank velocity fields, rather than the compressed data, although SHRED is trained on the latter. Behind this choice is the fact that our aim in application of SHRED to real-world flows is to accurately capture the full flow fields, not their compressed version. To explore the limitations on SHRED due to the compression step, we nevertheless include versions of these error metrics where the reconstructions are measured relative to the compressed data as “ground truth” in Appendix A.

1. Time-averaged velocity profile

The time-averaged velocity profile of a free-surface turbulent flow gives a sense of the effect of the upper boundary on the subsurface flow. It is essentially a measure of the root-mean-square velocity component(s) averaged over all snapshots, computed for component u by

$$\langle u_{\text{RMS}} \rangle = \left\langle \left(\frac{1}{mn} \sum_{x=1}^m \sum_{y=1}^n u_{x,y}^2 \right)^{\frac{1}{2}} \right\rangle, \quad (13)$$

where $\langle \dots \rangle$ denotes averaging in time (i.e., over all snapshots) and spatial averaging is performed as detailed in connection with Eq. (12). Unlike the metrics presented below, the time-averaged velocity profile is not a performance metric *per se*, but computing and comparing such profiles for ground truth data and reconstructed data enable a straightforward assessment of reconstruction performance.

2. Normalized mean squared error

Normalized mean squared error (NMSE) is one of the simplest and most widely used metrics for reconstruction accuracy. We use the normalized form:

$$\text{NMSE} = \frac{\langle \sum_{x=1}^m \sum_{y=1}^n |\tilde{u}_{x,y} - u_{x,y}|^2 \rangle}{\langle \sum_{x=1}^m \sum_{y=1}^n u_{x,y}^2 \rangle}, \quad (14)$$

where $\tilde{u}$ is the reconstruction field and u is the ground truth. Hence, the MSE is calculated as the spatial (over each plane) and temporal mean of the square of the difference between the ground truth and reconstruction, and normalized by the mean square of the ground truth. The MSE is a simple measure of the error in the values themselves, at either a single point or, as in this case, the averaged error of data points over a plane. The MSE captures the loss in amplitudes in the reconstruction. However, it has drawbacks such as not taking into account the sign of the signal, disregarding spatial structures, and treating each data point as equally important [63] as we discuss further below.

3. Power spectral density error

A common description of turbulence is to regard the power spectral density (PSD). Given snapshots $u(x, y, t_n)$, we define the 1D PSD for the single velocity component u along x , averaged over y and time, as

$$E(k) = \frac{1}{N_t N_y} \sum_{n=1}^{N_t} \sum_{j=1}^{N_y} |\mathcal{F}_x[w(x) u(x, y_j, t_n)](k)|^2, \quad (15)$$

where $\mathcal{F}_x$ denotes the discrete Fourier transform in x , N_t and N_y are the number of time steps and grid points in y direction respectively, and the Hanning window function is denoted by $w(x)$.

To compare the different flow cases, we make use of a normalized 1D spectrum, defined as

$$\text{NPSD}(k) = \frac{E(k)}{\max_k E(k)}, \quad (16)$$

i.e., spectra are normalized by the maximum of the ground truth 1D spectrum in each case.

In order to evaluate the reconstruction accuracy in terms of the 1D PSD of the velocity field u , we introduce the 1D power spectral density error (PSDE) as the relative error in integrated spectrum up to a cutoff wave number k_c (details below and in Table II), defined as

$$\text{PSDE} = \frac{|\int_0^{k_c} \tilde{E}(k) dk - \int_0^{k_c} E(k) dk|}{\int_0^{k_c} E(k) dk}, \quad (17)$$

where $\tilde{E}(k)$ and $E(k)$ are the 1D spectra for the reconstruction and ground truth data, respectively, computed by Eq. (15). In the discrete implementation we use Simpson's rule on the FFT wave number bins k_i with weights w_i , and the PSDE thus becomes

$$\text{PSDE} = \frac{|\sum_{i=0}^{i_c} w_i \tilde{E}_i - \sum_{i=0}^{i_c} w_i E_i|}{\sum_{i=0}^{i_c} w_i E_i}, \quad (18)$$

where i_c is the index for the wave number bin k_c .

We choose the cutoff wave number k_c to exclude scales that are not reliably represented by the low-rank data used in training. Operationally, k_c is set near the onset where the rank-truncated SVD spectrum begins to depart noticeably ($\approx 10\%$) from the ground-truth spectrum; in practice this falls in the intermediate (inertial-range) wave numbers for all datasets. The same k_c is used for ground truth and reconstructions within a given dataset. By setting these integration limits, rather than using the full spectrum, we neglect the major contribution in the error (when compared to the

full-rank ground truth) the low-rank SVD truncation process itself would produce. Therefore, it is only relevant to compare the spectral power of the SHRED reconstruction and full rank ground truth down to the low-rank SVD-resolved spatial scales. For the experimental data, this corresponds to a wavelength resolution of $\lambda_c = 2\pi/k_c = 1.5$ cm, which is about 20% (E1) and 13% (E2) of the integral length scales. For the DNS data, the cutoff wavelength is chosen at around 11% of the integral length scale.

4. Structural similarity index measure

The Structural Similarity Index Measure (SSIM) is a metric often used to compare image quality in terms of visual perception [64]. While MSE captures the point-by-point local mismatch of the average amplitude of an image, SSIM also emphasizes structural correlation and contrast, outperforming MSE and PSNR (see below) in evaluating for visual similarity as described, e.g., by Wang and Bovik [63]. It is designed to be more aligned with human visual perception, and is therefore often used in computer vision and deep learning. The SSIM we use is defined as

$$\text{SSIM} = l(\tilde{u}, u)c(\tilde{u}, u)s(\tilde{u}, u), \quad (19)$$

where $\tilde{u}$ is our reconstruction image, u is the full-rank ground truth, $l(\tilde{u}, u)$ is the luminance similarity factor, $c(\tilde{u}, u)$ the contrast similarity factor, and $s(\tilde{u}, u)$ is the structure similarity factor, all defined and discussed in [64]. The SSIM can take values between -1 (anticorrelated) and 1 (structurally identical), with 0 indicating no similarity.

5. Peak signal-to-noise ratio

The peak signal-to-noise ratio (PSNR) is a measure of image resolution widely used in computer vision and reconstruction tasks. Essentially, it quantifies how much of the signal in an image is relevant compared to irrelevant noise. It is inversely related to the MSE metric, and is defined as

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_j^2}{\text{NMSE}} \right), \quad (20)$$

where MAX_j is the maximum pixel bit value of image j , and NMSE is given as in Eq. (14). Generally, the more noisy and distorted the image, the lower the PSNR value. The metric is usually given in decibels (dB), and typical values for moderate quality images are 20–30, while exceptionally good quality images will show values of 30–50 [64]. Although it is closely related to the MSE, we include PSNR as a metric because it accounts for the maximal value in an image, creating a reference point for the signal-to-noise ratio.

IV. RESULTS AND DISCUSSION

In what follows, we present and discuss the results of reconstructing subsurface turbulent velocity fields from sparse measurements of the free surface only, by applying the SHRED algorithm to the turbulence data from the DNS [36] and experimental [37] datasets. We first show and discuss the field reconstructions for shallow and deep horizontal planes, and compare these to the uncompressed and compressed fields. Second, we present a detailed analysis of the depth-dependent performance of SHRED, using the set of error metrics presented in Sec. III C. Third, we compare the performance of SHRED with a proper orthogonal decomposition based method for sensor-to-flow mapping. Fourth, we demonstrate the capabilities of SHRED in temporal dynamics by looking at the RMS velocity time series of reconstructed fields. Finally, we show that the reconstructed fields yield reasonable turbulent power spectra for the relevant spatial scales, as compared to the compressed and ground-truth fields.

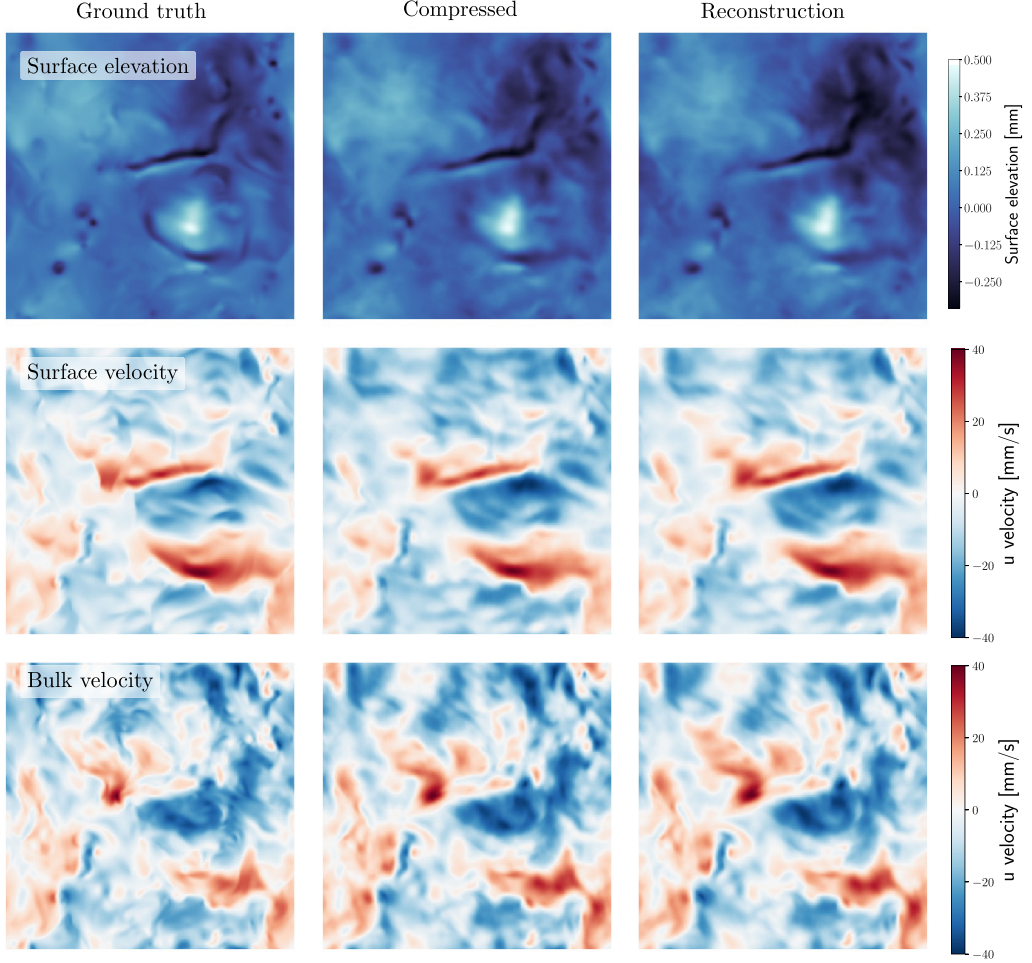


FIG. 6. Demonstration of the reconstruction capabilities of SHRED for DNS dataset S2: Ground truth (left-hand column), compressed training data with rank 250 (middle column), and SHRED reconstructions of the compressed data (right-hand column), for surface elevation (top) and horizontal velocity component u near the surface at depth $0.02L_\infty$ (middle row) and in the bulk at depth $1.2L_\infty$ (bottom). In the top row, light (dark) colors illustrate elevation (depression) of the surface compared to the mean level; in the middle and bottom rows darker red (blue) indicates higher positive (negative) values of the velocity u with white representing zero velocity. The size of each plotted region is approximately $6L_\infty \times 6L_\infty$.

A. Reconstruction of surface elevation and velocity fields

To illustrate the performance of SHRED qualitatively, we present a side-by-side comparison of the surface elevation and a component of the horizontal velocity field: the original, uncompressed data, their low-rank SVD approximations, and the SHRED reconstructions, for the S2 DNS data in Fig. 6 and the experimental dataset E2 in Fig. 7. The immediate eyeball observation is that the fields lose some sharpness in the SVD compression step, whereas the compressed and reconstructed fields are only distinguishable by eye upon careful inspection.

From Fig. 6 one might get the impression that errors in the final reconstruction of the fields (right-hand column) occur during compression before training while the subsequent steps reproduce the compressed fields near-perfectly. To some extent, this is correct. However, increasing the

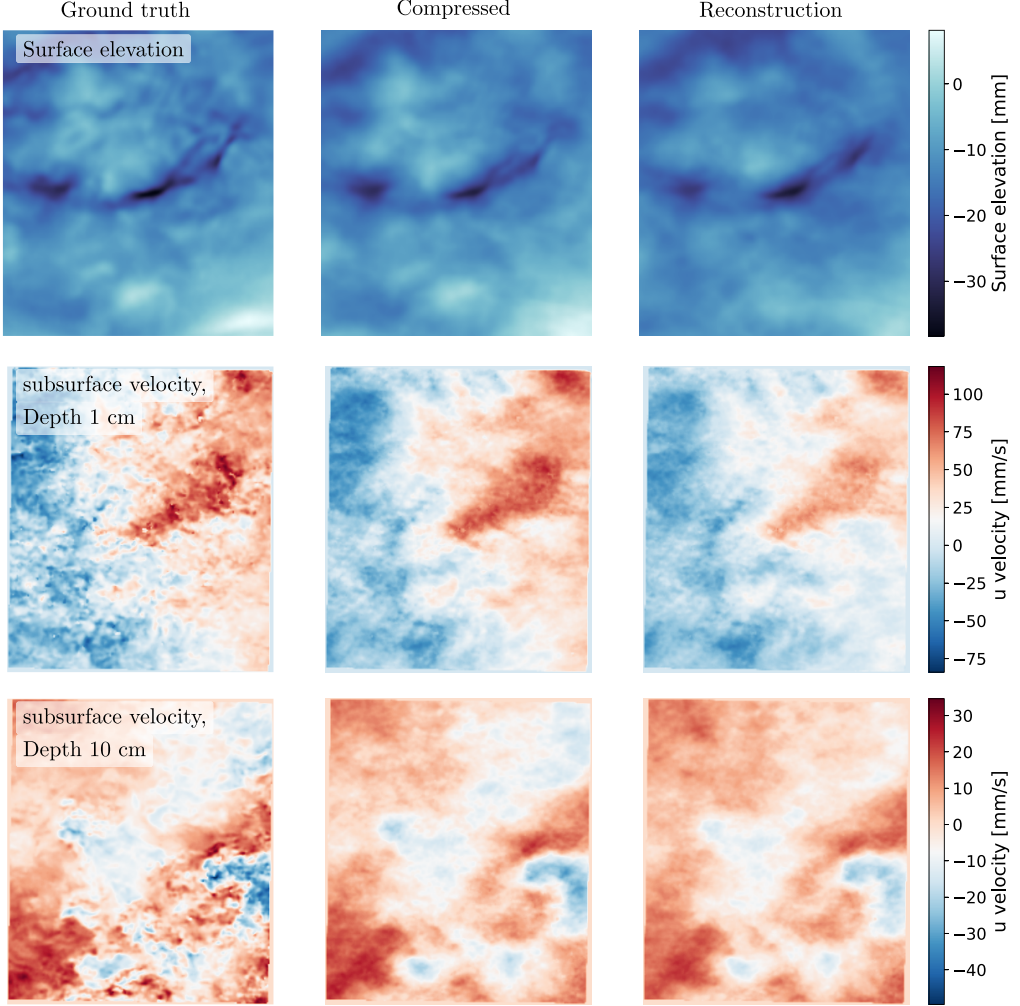


FIG. 7. Same as Fig. 6 but for experimental data, flow case E2: Ground truth (left-hand column) compared to compressed training data of rank 100 (middle column) and SHRED reconstructions of the compressed data (right-hand column), for surface elevation (top) and the corresponding horizontal velocity component u , 1 cm ($0.10L_\infty$) and 10 cm ($0.85L_\infty$) beneath the surface, in the middle and bottom rows, respectively. The surface elevation is simultaneous with the velocity in the middle row, whereas measurements in the deeper plane were taken at a different time; their corresponding surface elevation is not shown here. The size of the surface elevation field is $1.86L_\infty \times 2.1L_\infty$, and slightly smaller at $1.68L_\infty \times 1.92L_\infty$ for the velocity fields.

number of SVD modes in compression beyond the ranks chosen does not necessarily improve the reconstruction. As we detail in Appendix A, training SHRED using the original uncompressed dataset or with very little compression leads to considerable overfitting and far poorer results, as well as higher computational cost.

Inspecting the surface elevation for DNS S2 data in the top row of Fig. 6, we notice a sharp, dark, depression curve—referred to as a “scar” [1]—and brighter areas which are “boils” which signify upwelling of fluid to the surface. These features are successfully reconstructed, although the amplitudes can be seen to have been somewhat dampened. Scars also manifest in the surface velocity field (middle row) as areas of fast flow. As with the surface elevation, the reconstruction

looks visually indistinguishable from the low-rank approximation and the ground truth. Moreover, as demonstrated in the bottom row in Fig. 6, SHRED is capable of reconstructing the velocity field in the bulk flow, that is, at depths below the surface-influenced layer (depth $\bar{z} \gtrsim L_\infty$ for dataset S2; details on the surface-influenced layer are in [2]).

Similarly, for the experimental dataset E2, Fig. 7 shows visually good reconstruction performances across several planes. It is a considerably more challenging task to reconstruct planes in the experimental flows due to noise and the much greater disparity of time and length scales at the far higher turbulent Reynolds numbers. However, we observe that even in this case, SHRED is capable of reconstructing from the time series of surface elevation in only three points, large and strong features such as scars at the surface elevation and medium-sized structures in the velocity fields as far down as 10 cm ($z \approx L_\infty$) below the surface, although smaller features are somewhat blurred. More clearly than in the DNS flow in Fig. 6, velocity amplitudes are generally diminished in the reconstruction (right column) compared to the compressed training data (middle column).

Figures 6 and 7 illustrate that SHRED is capable of qualitatively reconstructing flow fields from sparse height measurements at the free surface in free-surface turbulence data from DNS as well as state-of-the-art experiments. SHRED’s performance is somewhat weaker in the experimental case, which is to be expected due to the presence of noise, greater range of turbulent scales, and significantly less data available for training compared to the long, continuous DNS datasets. In light of this we find it remarkable that SHRED reconstructs features of centimeter size 10 cm beneath the surface, well outside the “blockage layer” where the free surface directly influences the velocity field [2,65].

B. Depth-dependent performance

To quantify the performance of SHRED beyond simple comparisons, we use the five metrics discussed in Sec. III C to evaluate different aspects of SHRED’s reconstruction of the horizontal velocity field u . As stated in Sec. III C, we report all error metrics relative to the original, uncompressed fields unless otherwise stated. The primary objective is to reconstruct the true flow dynamics, with compression serving solely as an intermediate representation to facilitate optimal learning. A detailed analysis of the effect of the SVD compression on the error is provided in Appendix A. In addition to the averages taken in horizontal space and time, we perform ensemble averaging over 25 individual reconstructions for each flow case, where every time series in the ensemble is constructed by a random distribution of the full dataset into training, validation, and testing. The results are displayed with depth z scaled by the integral length scale on the ordinate axis. Note that we adopt the convention from Aarnes *et al.* [2] of using an average measure for “horizontal” grid plane depth for the DNS data, to allow for straightforward comparison of flow variables at a grid plane on the undulating grid without interpolation (details are found in [2]).

Because the error metrics are averaged over time and ensembles, which could potentially conceal the direct performance error image by image, we first consider instantaneous profiles of $u_{\text{RMS}}(z, t)$ in Fig. 8, comparing reconstructions (dashed lines) to the uncompressed ground-truth values (solid lines) at three different arbitrarily chosen time instants. Note that for the two experimental flows, the subsurface velocity measurements taken at different depths z are not taken at the same time since different planes were measured separately. Generally, the reconstruction profiles match the ground truth profiles in shape but for a drop in magnitude, different for different depths and instants. There is a tendency for rapid changes in the measured profiles to be smoothed out in the reconstructed ones, particularly visible here for flow S1 at instant 1. Such rapid changes are typically caused by intermittent turbulent events which are difficult for a neural network to learn, in particular for time-series-based networks such as the LSTM method used in SHRED. Strikingly, in all examples the accuracy of reconstruction deteriorates only very slightly with increasing depth.

Shifting our attention to time-averaged results, the corresponding time-averaged quantities, $\langle u_{\text{RMS}}(z, t) \rangle$, are plotted in Figs. 9(a) and 9(b) for DNS and experimental data, respectively, after

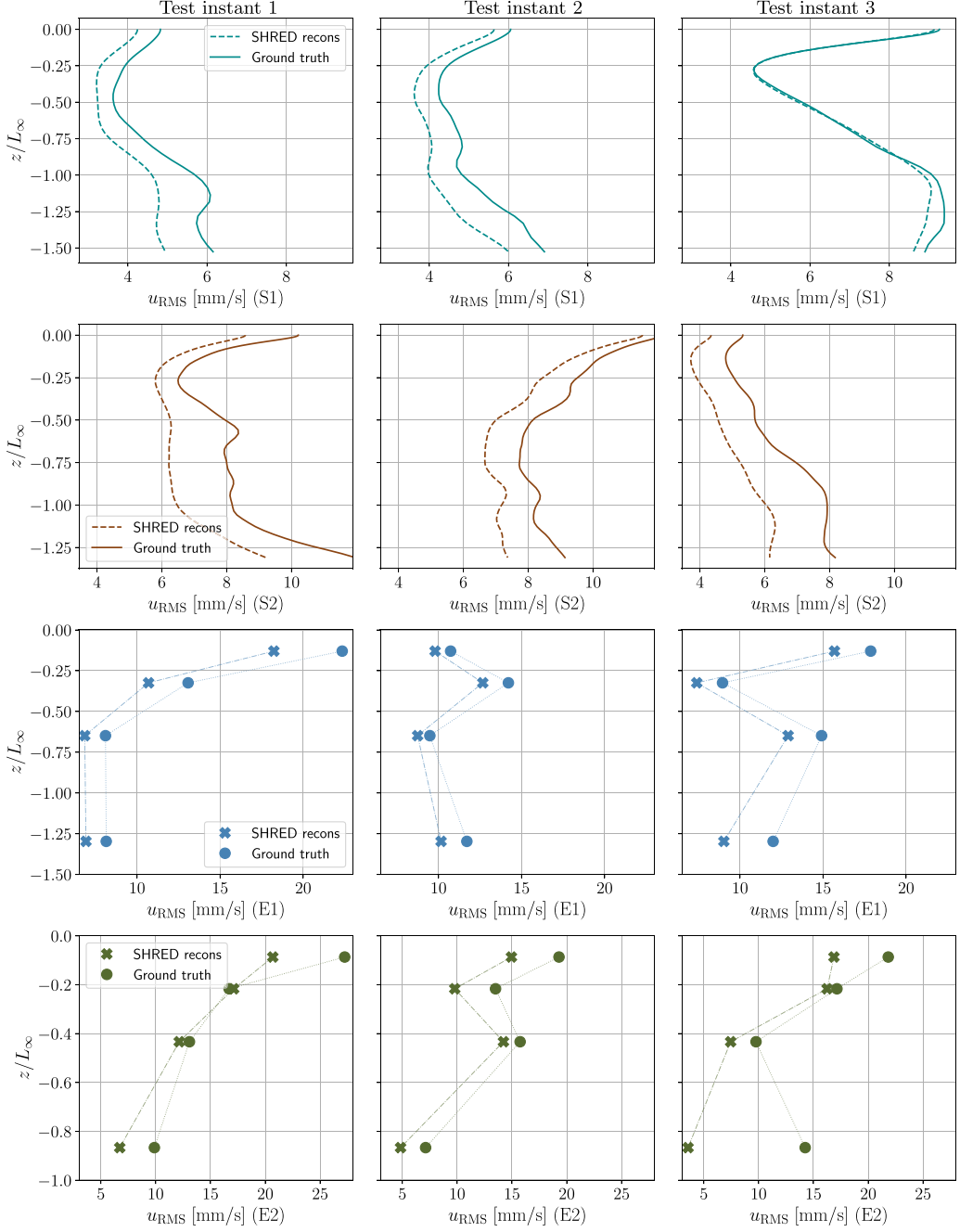


FIG. 8. Examples of instantaneous profiles of $u_{\text{RMS}}(z, t)$ for ground truth (solid lines) and reconstruction (dashed lines) for the four datasets S1, S2, E1, and E2 (rows from top to bottom) at three randomly chosen time instants (columns from left to right). Note that for S1 and S2, the vertical profile per dataset and test instant is simultaneous, while for E1 and E2 the u_{RMS} at each plane is measured independently from other planes, i.e., vertical profile is not simultaneous, but rather a combination of four separate instants, each with their separate SHRED model.

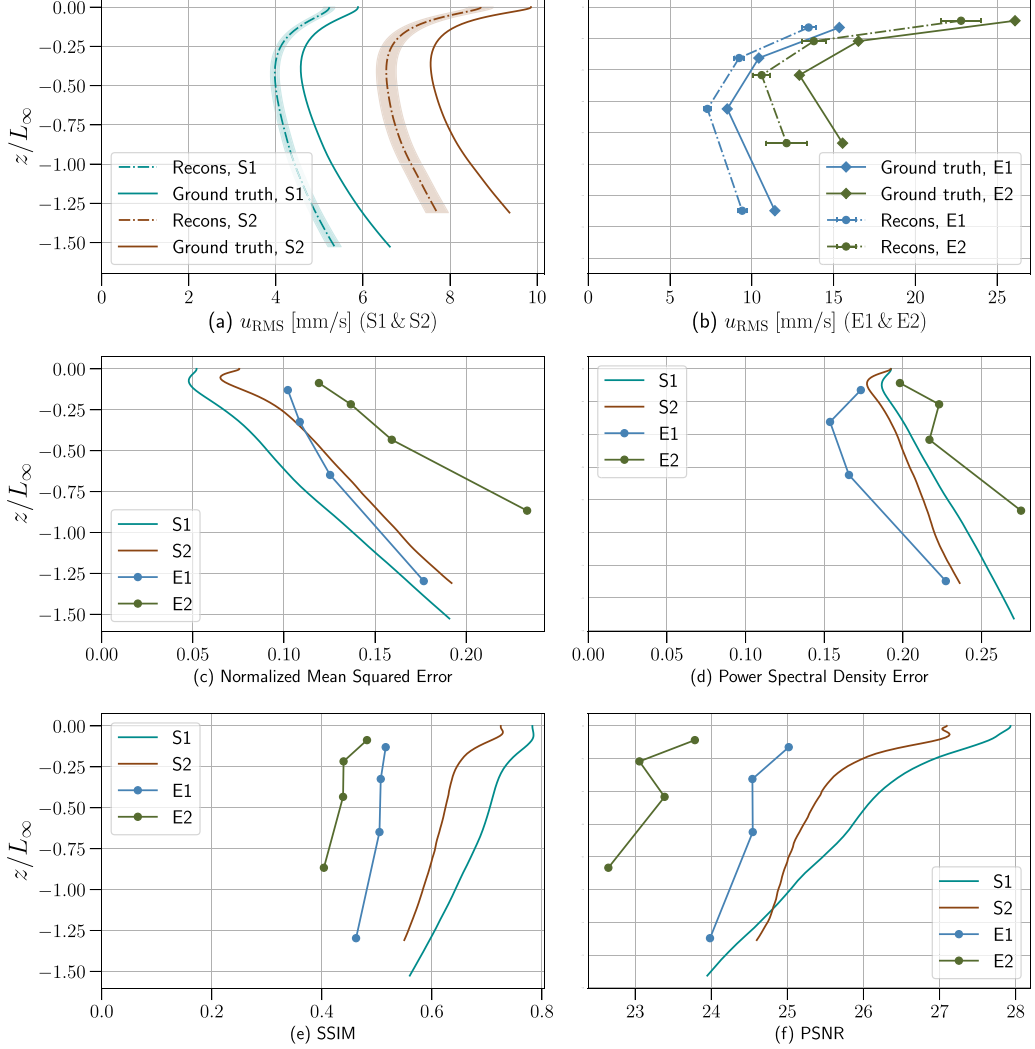


FIG. 9. Comparison of error metrics at different depths beneath the surface for the four datasets in Table I, taking the ensemble average of 25 independent SHRED reconstructions. Panels (a) and (b) show reconstructed values of $\langle u_{RMS} \rangle$ (dashed lines) for DNS and experimental data, respectively, compared with ground truth measurements (solid lines). The standard deviation of the ensembles are shown as shaded areas (a) and error bars (b). The bottom four panels show error metrics NMSE (c), PSDE (d), SSIM (e), and PSNR (f), all defined in Sec. III C. All errors are measured relative to the uncompressed fields.

averaging over an ensemble of 25 individual SHRED reconstructions each using a different set of three randomly chosen surface points as their sparse input. By normalizing the depth coordinate by the integral scale L_∞ for each flow, the profiles can be compared side by side. Again, the solid lines represent the uncompressed data (i.e., ground truth), and color-matching dashed lines mark their reconstructed counterparts; the standard deviation of the ensemble is shown as shading [Fig. 9(a)] or error bars [Fig. 9(b)]. As was observed at individual instants in Fig. 8, the trends of the profiles are retained by the reconstruction, but consistently offset to lower values, corresponding to a loss of energy in the reconstruction. This can only partly be explained by the compression which truncates the number of SVD modes; the offset is larger than expected by a rank reduction alone, as using a

moderate truncation value for r has a small impact on the large, energy-carrying scales of the flow (as seen in Fig. 4). For an in-depth discussion of the effect of data compression on the results from SHRED, see Appendix A.

The depth variations of the normalized mean squared error (NMSE) of Eq. (14) are shown in Fig. 9(c). For the DNS flows we notice a low MSE of about 6%–7% close to the surface, and a steady increase up to around 19% for the lowermost planes. These are generally significantly lower values than the subsurface reconstruction MSE of the CNN method of Xuan and Shen [22] achieved, although it should be noted that this comparison is not a direct one-to-one benchmark due to differences in configurations. Differences aside, we note that both SHRED and the CNN-based method proposed in [22] tend to result in dampened magnitudes of large-scale turbulence, for the surface-based flow reconstructions. That the NMSE is higher in the lower planes might be expected since there is hardly any correlation in a direct sense between the surface motion and the flow field at this depth.

We observe that the 1D PSD error metric (PSDE) follows similar trends as the NMSE. The errors are generally in the range of 15%–30% in the large-to-intermediate length scales. The reconstructions of the DNS data show similar power spectral errors, although the high-Re dataset S2 performs slightly better by this metric. The experimental flow E1 shows lower errors than the DNS cases, which is also reflected in the PSD spectra of Fig. 11. This is, perhaps, unexpected because the data are generally noisier and harder to reconstruct, as highlighted by the other metrics. In contrast, the second experimental flow, E2, has the largest PSD error. We shall see below that for all datasets the PSD error is always associated with a loss in turbulent kinetic energy in the reconstructions, at all lengthscales. Other turbulence-sensing reconstructions find similar losses in turbulent kinetic energy (see [20,22]), although these are hard to compare directly with our results due to different flow conditions and selection of error metrics.

The SSIM results are presented in 9(e). The SSIM metric differs from the other error metrics that we use, as it targets structure, luminosity, and contrast, while the physical interpretation of the metric is not as obvious. The SSIM results reflect the qualitative observations made from Figs. 6 and 7, that there is an indisputable visual similarity between reconstructed fields and ground truth when the larger structures of the turbulent fields are considered. The DNS reconstructions yield SSIM values between 0.7 and 0.8 near the surface, falling to between 0.55 and 0.65 at one integral length scale depth. For the experimental flows, peak values are found near the surface at around 0.47 to 0.52, falling to around 0.4 to 0.45 in the deepest planes. The latter might be considered a weak result in general image reconstruction schemes, not so in a turbulence context, considering that it is based on the time dynamics of three surface points only, after training. The SSIM likely emphasizes small-scale structures of the full-rank ground truth which are no longer present in the compressed training data (or, if they were, would be subject to overfitting). This is particularly prominent for the SSIM results for experimental data. For the DNS data, the reconstruction SSIM of almost 0.8 near the surface based on very sparse measurements, can be compared to super-resolution schemes for reconstructing full-state space from a coarse-grained and sparse field. The results of super-resolution reconstructions differ from case to case, but a study based on data from turbulent large-eddy simulation (LES) found many methods to produce values of 0.8–0.9 [66]. In this regard, SHRED is close to matching this performance while also producing acceptable SSIM results for planes far from the surface.

For the SHRED reconstructions, it is clear that the PSNR values, as seen in Fig. 9(f), are within acceptable values (above 20 dB) for all datasets. We are not aware of other sufficiently similar reconstruction studies to which the values can be directly compared, yet one might note that super-resolution reconstructions of turbulent DNS data have found similar PSNR values [10]. The performance is matched, even with planes as deep as L_∞ or further from the surface, with PSNR values of 22.5–25.5 dB. The DNS datasets yield higher values than the experimental ones, whereas the more turbulent ones show lower values. As expected, the PSNR value generally decreases with depth, with the largest decrease occurring near the surface in all cases, where the flow changes rapidly due to surface viscous and blockage effects from the surface [2,65].

TABLE III. Energy retained in the reconstructed velocity fields, relative to the energy in the ground truth data. Computation as a cumulative measure of energy retained from the surface down to a fixed depth, according to Eq. (21).

Dataset	$E(z \leq 0.1 L_\infty) (\%)$	$E(z \leq 0.5 L_\infty) (\%)$	$E(z \leq 1.0 L_\infty) (\%)$
S1	79.0	78.3	76.9
S2	78.7	77.6	75.8
E1	75.9	75.3	73.7
E2	72.6	71.3	69.6

The dampened magnitudes in the SHRED reconstructions, which we observe in the the depth profiles of u_{RMS} in Figs. 8 and 9(a)–9(b), indicate a systematic energy loss. We quantify how much energy SHRED is able to retain in the reconstructions, by using $\langle u_{\text{RMS}}^2(z_k) \rangle$ as a single-component measure of the kinetic energy of a horizontal velocity plane at depth z_k . The relative energy retained from the surface to a selected depth z_l is then computed as a cumulative measure by

$$E(z \leq z_l) = \frac{\sum_{z_k=z_0}^{z_k=z_l} \langle \tilde{u}_{\text{RMS}}^2(z_k) \rangle}{\sum_{z_k=z_0}^{z_k=z_l} \langle u_{\text{RMS}}^2(z_k) \rangle}, \quad (21)$$

where $\tilde{u}$ and u denote the reconstruction field and ground truth, as above, and z_0 denotes the position of the uppermost plane for velocity measurements in each dataset (coinciding with the surface for DNS datasets, but not for experiments).

The relative energy retained down to three different depths is reported in Table III. Note that as the experimental datasets E1 and E2 are sparse in depth, calculation down to $z_l = 0.1 L_\infty$ includes only the top plane, $z_l = 0.5 L_\infty$ includes the three uppermost planes, and $z_l = 1.0 L_\infty$ includes all four planes. We observe that the retained energy decreases gradually with depth for all datasets, consistent with the increase of errors and reduction of surface-to-subsurface correlation with depth, with roughly 70%–80% of the energy captured by the SHRED reconstructions down to one integral length scale below the surface. Although one might expect that much of the energy loss is due to the fact that SHRED exclusively trains and validates on a significantly compressed dataset, the compression itself only accounts for 2–7 percentage points of the energy loss (details in Appendix A). A glance at Fig. 11, which we consider in detail later, indicates that it is SHRED’s energy loss at the large scales (low wave numbers) which is the major driver of the total energy loss. The effects noted here may be tied to the problem of properly representing the large, highly intermittent upwelling events in our free-surface flow. The very largest of these are rare, e.g., occurring only 5–10 times in each of the DNS datasets, making it very hard to fully account for them across our training data. The LSTM, despite good temporal performance, is best suited for learning smooth dynamics and together with the decoder, they might underestimate intermittent magnitudes of energetic scales.

Overall, the depth-dependent error metrics indicate that even 1.0–1.5 integral length scales from the surface, large-to-intermediate-scale turbulence can be reconstructed well enough for many practical purposes from just three measurement points of the surface elevation, within the time range of training data. As a proof-of-concept study these results demonstrate the potential of SHRED for remote sensing applications, where only observations *at* the surface, and not beneath it, are available. The ability to reconstruct bulk flow structures using just three surface points demonstrates a key step toward remote sensing of subsurface turbulence. However, we emphasize that although the time steps where reconstruction is performed are not part of the actual training set, they lie within the same time range used in training. Future work will explore the capability of SHRED to make reconstructions of previously unseen flow regimes or truly independent test cases, which is crucial for generalizability and real-world deployment.

TABLE IV. Performance comparison of SHRED and POD-based linear regression at multiple depths for datasets S2 and E2.

Dataset	Depth	Metric	SHRED	POD + linear regression		
			3 sensors	3 sensors	30 sensors	300 sensors
S2	$z/L_\infty \approx 0.1$	NMSE	0.072	1.028	1.861	0.029
		SSIM	0.704	0.043	0.124	0.775
	$z/L_\infty \approx 0.5$	NMSE	0.123	1.028	1.749	0.070
		SSIM	0.625	0.042	0.100	0.688
	$z/L_\infty \approx 1.0$	NMSE	0.161	1.020	1.786	0.451
		SSIM	0.582	0.048	0.105	0.593
E2	$z/L_\infty \approx 0.1$	NMSE	0.124	1.527	0.584	0.706
		SSIM	0.486	0.048	0.316	0.405
	$z/L_\infty \approx 0.5$	NMSE	0.158	1.276	0.284	0.309
		SSIM	0.432	0.012	0.356	0.335
	$z/L_\infty \approx 1.0$	NMSE	0.240	1.252	0.440	0.422
		SSIM	0.402	0.031	0.267	0.296

C. Comparison with linear, POD-based method

To evaluate SHRED as an overall sensing and reconstructing method, we compare its performance with that of proper orthogonal decomposition (POD) of fields with linear regression estimation for sensors-to-POD mapping. The linear regression estimation uses least-squares regression mapping from lagged sensor measurements to r number of POD coefficients, where we set r equal to the low-rank values from Table II. We perform the comparison on the the highest Reynolds number dataset from the DNS and experiments, datasets S2 and E2, respectively, and compute the NMSE and SSIM for multiple depths for each dataset. As in Sec. IV B, we report ensemble-averaged error metrics, with 25 ensemble runs used for the SHRED computations and 10 for the POD with linear reconstruction. The results are presented in Table IV.

While POD provides an efficient low-dimensional representation of the flow, the results in Table IV indicate that recovering its coefficients from linear mapping of sparse surface measurements is highly challenging. With the identical three sensor setup that we have so far used for SHRED reconstructions, the POD-based linear regression mapping performed poorly, with NMSE above 1 and SSIM close to zero, indicating that errors exceeds signal energy and that spatial structures are not well preserved in the output. In short, for the identical sparse-sensor configuration as SHRED, the POD-based linear regression method comes nowhere near SHRED's performance.

Additional tests show that POD sensing performance depends strongly on the number of surface sensors, whereas the dependence on retained rank is weak. We therefore include POD sensing results for configurations with 30 and 300 sensors to level the playing field between the two methods we compare here. As the POD-based regression method is linear, performance generally improves with more inputs. SHRED, on the other hand, is nonlinear, and its performance improves with sensor count only to a certain point, while high input dimensions increase complexity of the learning and degrades reconstructions (details in Appendix B). From the POD-regression analysis, we find that approximately two orders of magnitude more input sensors are needed to match the ultrasparse three-sensor SHRED performance for our datasets. However, even with more sensors, the magnitude of deeper velocity planes (here $z/L_\infty \approx 1$), are poorly captured by the POD-based linear regression. Hence, although the POD-based linear method yields comparable results to SHRED for dense inputs, SHRED outperforms it substantially in the sparse-sensor regimes, as well as for surface-to-depth sensing, and is thus more powerful for turbulence sensing of limited

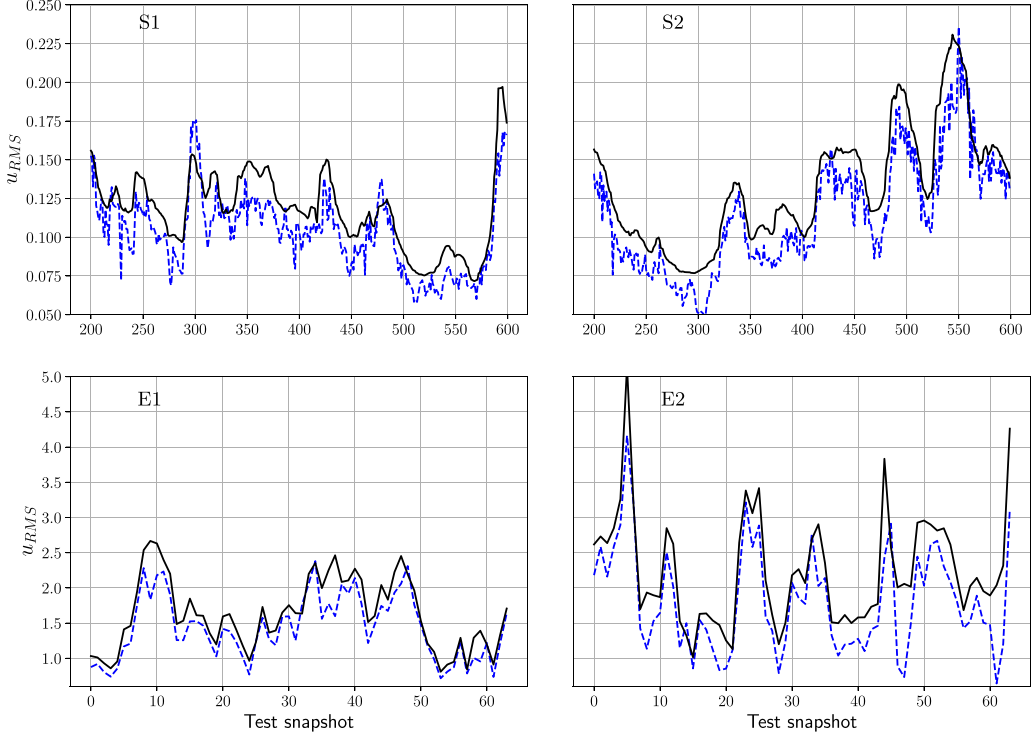


FIG. 10. Examples of ground truth (black solid line) and reconstruction (blue dashed line) of single-frame rms velocity, u_{RMS} , a time series of test snapshots near the surface for all four datasets in Table I at depths of around $0.08\text{--}0.12L_\infty$.

data. These findings are consistent with previous studies demonstrating nonlinear decoder networks outperforming linear methods in sparse regimes [8], as well as the findings by Williams *et al.* [28], showing SHRED outperforms linear POD-based methods for a range of high-dimensional spatiotemporal datasets, including isotropic turbulence.

D. Temporal analysis of planar root-mean-square velocity

As an illustration of the SHRED reconstruction capabilities of temporal dynamics, we show the time series of the planar RMS values for a single SHRED reconstruction case, in Fig. 10. The planes chosen are relatively close to the surface, 1 cm depth for the experimental flows ($0.21L_\infty$ for E1 and $0.14L_\infty$ for E2), and planes at $0.2L_\infty$ depths for DNS flow cases S1 and S2. We observe a high degree of correlation between the time series of the ground truth and the reconstruction. The normalized cross-correlation values with zero lag, are above 0.94 for all datasets except E2, in which the correlation value is 0.89. The values of u_{RMS} are generally lower for the reconstructed fields than the ground truth, indicating some loss of kinetic energy, as discussed in Sec. IV B. Notably, the error in u_{RMS} does not increase dramatically when intermittent high-intensity turbulent events, corresponding to sudden peaks in the RMS velocity, occur.

E. Spectral analysis

In this section, we investigate how the power density spectra of SHRED-reconstructed fields perform compared to uncompressed data (ground truth) and the compressed, low-rank data. An example is shown in Fig. 11, for reconstruction of flow at a depth of approximately L_∞ . The 1D

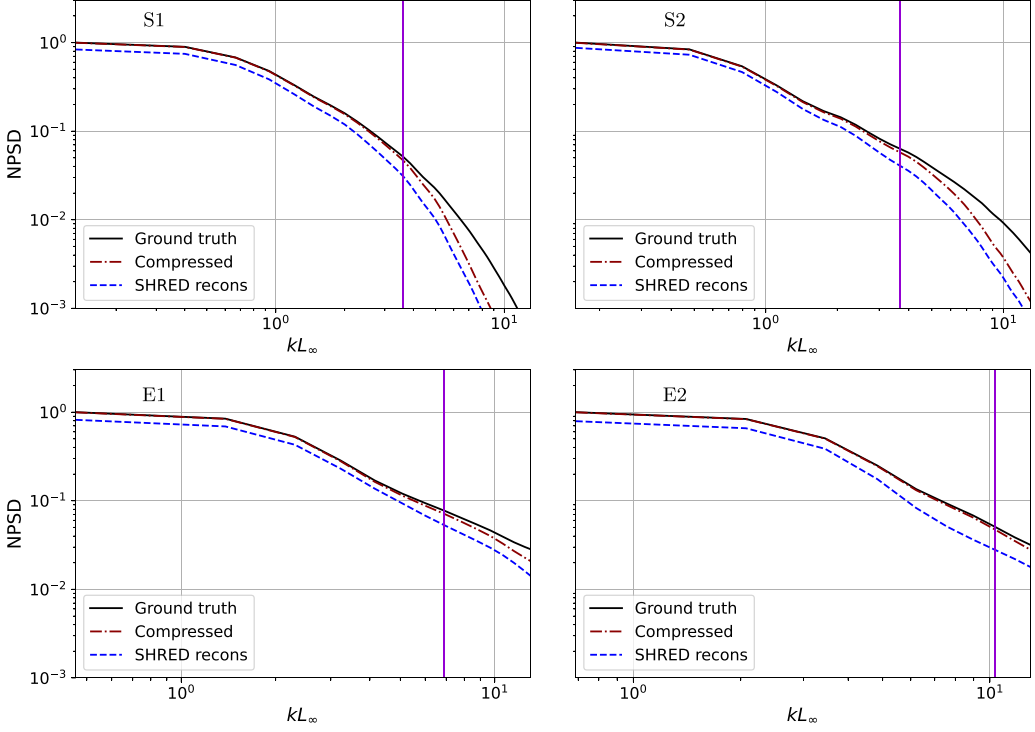


FIG. 11. Normalized 1D power spectral density (NPSD) for the four datasets, including results for the uncompressed, ground truth data (dark, solid line), low-rank compressed data (dash-dotted, red line), and SHRED reconstructions (blue, dashed line), for the velocity fields at depths $\approx 1.0L_\infty$. The purple vertical line indicate the cutoff wave number for the SVD compression, i.e., the highest resolved wave number in the velocity fields of the compressed data and SHRED reconstructions. The spectra are normalized by the maximal value of the ground truth spectra, and the wave number is normalized by the inverse integral length scales in each case.

power density spectrum in the x direction, averaged over time and spatially in the y direction, is calculated in each case. The purple dotted vertical line indicates the cutoff wavenumber for the low-rank truncation, i.e., the wave number at which the compressed spectrum deviates from ground truth by about 10%. (See Sec. III B and Appendix A for details on selecting the rank truncation.) One observes that generally, the spectra for the compressed field (dashed dotted red line) closely follow the spectra of the ground truth (solid dark line) up to the cutoff wave number. We observe that the spectra of the SHRED reconstructions generally follow the ground truth, but with a spectral energy loss that is fairly consistent in log-space over all frequencies. However, there are somewhat larger relative errors in the high-frequency end, suggesting fine-scale dynamics are more challenging to capture. On the other hand, the majority of the total kinetic energy is contained in the lower wave numbers, hence the low-frequency spectral errors contribute the most to the energy loss.

V. CONCLUSIONS

In this study, we demonstrated a proof-of-concept application of the SHallow REcurrent Decoder (SHRED) for reconstructing subsurface turbulence fields from sparse surface measurements in free-surface flows. Using only three arbitrarily placed sensors measuring surface elevation as input, SHRED was able to infer the dominant large- and intermediate-scale structures below the

surface across four different datasets of turbulent flow beneath a free surface together with the instantaneous elevation of the surface itself, two based on simulation data and two on experimental data recently published in [2,38]. The differences between the datasets in terms of Reynolds number, sparsity, and noise demonstrate the flexibility of SHRED in handling different turbulence regimes.

The results show that SHRED preserves the most energetically significant flow features, and performs particularly well near the surface, even on highly multiscale experimental data. While reconstruction accuracy decreases with depth, SHRED still provides meaningful results as deep as one integral length scale below the surface. Importantly, this was achieved without direct access to any subsurface measurements at inference time, highlighting the potential of SHRED as a tool for remote sensing of subsurface turbulence from very sparse surface measurements. Note, however, that subsurface fields are necessary during the training step, limiting SHRED to datasets which have (at least partially) temporally resolved flow fields below the surface.

This work addresses the central challenge of estimating near-surface turbulence in rivers and oceans from surface observations alone. Such capability is crucial for quantifying gas and heat fluxes at the water-air interface, where *in situ* measurements are impractical at scale. The demonstrated ability of SHRED to learn nonlinear mappings from sparse input to high-dimensional turbulent states marks a step forward toward scalable, nonintrusive field sensing.

Future work should aim to improve generalization across flow regimes, as the current validation setup is limited to data drawn from the same underlying datasets. Moreover, recent research [53] has demonstrated that SHRED has potential for forecasting in time, and demonstrating this for the complex case of free-surface turbulence would be most advantageous. Due to the intermittency of the largest structures we observe in our datasets a longer time series than we have available is necessary to achieve full flow field forecasting, and we have limited the present study to spatial reconstruction and temporal in-betweening. Pairing SHRED with other methods, like in SINDy-SHRED [35], which combines sparse sensing and reconstruction with identification of nonlinear dynamics, or SENDAI [67], which includes learned residual corrections across spatial scales, could potentially improve the reconstruction accuracy outside of the training time domain and could help make the model generalizable to real-world flows. Other possible steps towards remote sensing could be to extend the model to reconstruct derived quantities such as energy fluxes or gas exchange rates. Ultimately, SHRED offers a foundation for machine learning-based frameworks for remote sensing, and opens a path toward real-world applications in oceanography, river monitoring, and environmental sensing.

ACKNOWLEDGMENTS

The experimental datasets are from an experiment performed by Ali Semati, Dr. Amélie Ferran, and Dr. Yi Hui Tee under the guidance of Prof. R. Jason Hearst; beyond the experimental data itself we have benefited from discussions with these colleagues and input on the manuscript. DNS data were generously shared with us by Prof. Lian Shen and Dr. Anxing Xuan at the University of Minnesota. We thank Omer M. Babiker for assistance with data handling and preparation and many discussions. The work of J.R.A. and S.Å.E. was cofunded by the Research Council of Norway (iMOD, Grant No. 325114) and the European Union (ERC CoG, WaTurSheD, Grant No. 101045299). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The work of J.N.K. was supported in part by the U.S. National Science Foundation (NSF) AI Institute for Dynamical Systems, Grant No. 2112085. J.N.K. further acknowledges support from the Air Force Office of Scientific Research (FA9550-24-1-0141).

K.S.M. and J.R.A. performed the data analysis with SHRED and wrote the first draft of the majority of the manuscript. J.N.K. made the first implementation, codeveloped the concept, and

contributed writing and supervision. S.Å.E. contributed to the concept and discussions and supervision. All authors contributed significantly to the writing of the final manuscript.

DATA AVAILABILITY

All code used in producing these results are included and thoroughly presented at [54]. The DNS data and supporting codes are available at [36]; the experimental data are available at [37].

APPENDIX A: EFFECT OF SVD COMPRESSION ON SHRED PERFORMANCE

As discussed in the main text, we compress the datasets before training and validation of SHRED. The compression is done by truncating the SVD representation of the flow to a fixed rank r , hence, the compressed dataset contains only the r most energetic SVD modes. The choice of rank has a significant impact on the reconstruction performance of SHRED. If the value of r is too low, relevant small-scale turbulence might be lost. If too high, too much noise might be included, with a risk of overfitting. In the present section, we include details on the selection of rank for all four datasets and investigate the effect of data compression on the depth-dependent error metrics and energy retention which we considered in Sec. IV B.

1. Parametric study of optimal choice of rank

From Fig. 12 we observe that the 1D PSD spectra for the SHRED-reconstructed fields (full colored lines) generally deviates from both the compressed (dashed-star lines) and the ground truth (black solid line) for a range of compression levels. Specifically, in the energy-containing region at low wavenumbers, the compressed spectra follow ground truth closely for all but the very highest compression level (lowest rank, $r = 25$), whereas the reconstructed field spectra appear to deviate further from ground truth for the higher rank values (right panel). We take this as an indication that it is SHRED itself, not the SVD compression, that contributes the most to the spectral error of reconstructed fields for the large-to-intermediate scales. We return to this point in Appendix A 2.

To quantify the performance of SHRED as a function of rank used in the data compression step, the selected error metrics are averaged depthwise, and calculated for a range of rank truncation values. Figure 13 shows the rank dependency of the SHRED error metrics for all four datasets. The

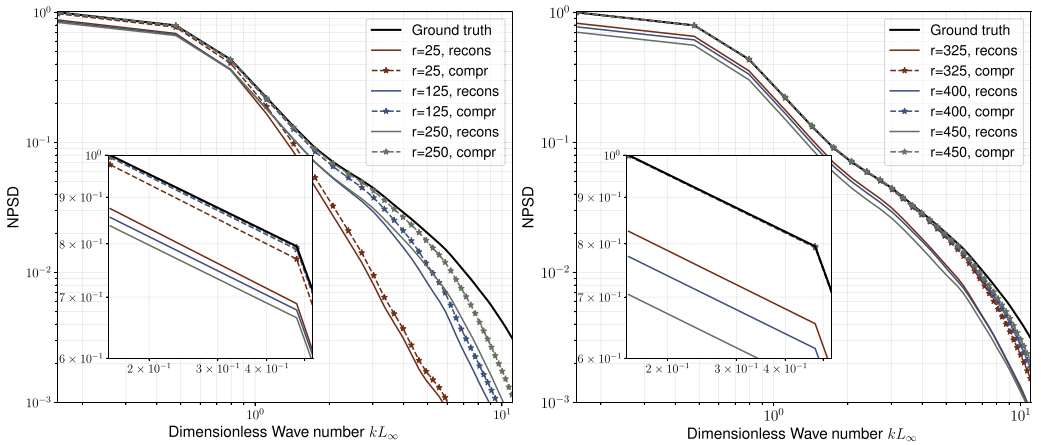


FIG. 12. Normalized 1D power spectral density (NPSD) of the low- and moderate-rank value r (left) and high-rank value (right) for dataset S2. Solid lines represent data reconstructed by the SHRED model, dashed-starred lines compressed data, and the solid black line the ground truth spectrum. Inset: Zoomed in view of the largest scales.

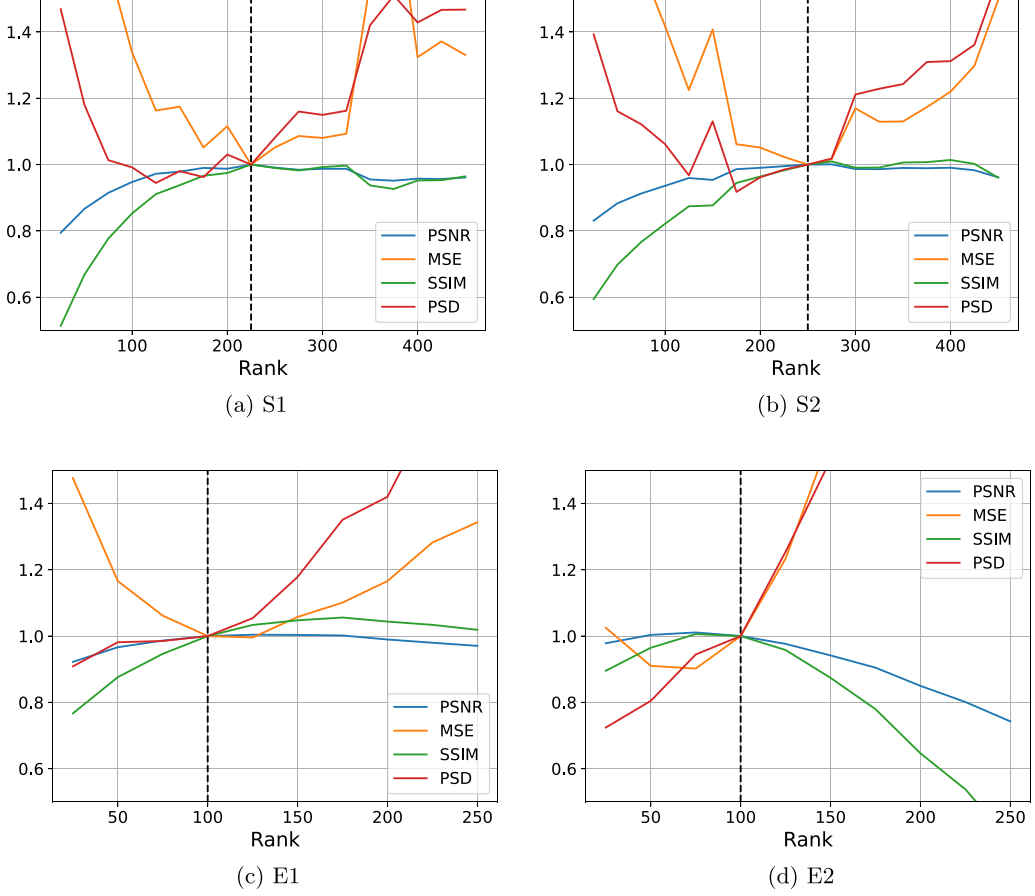


FIG. 13. Normalized error metrics averaged in depth for different levels of data compression, where higher rank denotes less compression. The dashed line marks the rank used throughout the study (see Table II). The error metrics are normalized relative to the mean error at this rank for each dataset. Note that optimal SHRED performance is met if PSNR and SSIM are maximal, while the same is true when PSD and MSE are minimal.

metrics are normalized by the error values at the rank we use for our main results, i.e., $r = 225$ for S1, $r = 250$ for S2 and $r = 100$ for E1 and E2 for scaling and comparison purposes. This rank value is marked by the dashed black lines. The range of the rank values we test is different between simulation and experimental datasets, due to the difference in total number of modes: the full rank of the experiments is $r = 900$, while for the DNS it is 10 900 for S1 and 12 500 for S2.

The error metrics indicate that the PSNR and SSIM values increase up until a certain rank number, while PSD and MSE generally decreases to a minimal error before increasing rapidly if r is further increased. This indicates that for too low rank numbers, there is a significant lack of information left in the low-rank representation, such that although SHRED might perform better in reconstructing the compressed input fields, the SVD itself has left out a significant amount of information. Hence, the fields lack structure and contrast for high SSIM values, and amplitudes and energy content of the largest modes are low, making NMSE and PSDE high. On the other hand, the error metrics generally show that SHRED performance decreases if the rank number is too high. In this case, the low-rank representation includes more fine details, including some noise. The temporal dynamics of the structures on these scales is highly intermittent and random, hence notably harder to reconstruct, especially from surface dynamics only. In between these two regimes,

TABLE V. Retained energy measure of the reconstructed fields, relative to the uncompressed and compressed data, respectively. All numerical values are in %.

Dataset	$z \leq 1.0 L_\infty$	
	Uncompressed	Compressed
S1	76.9	79.8
S2	75.8	80.7
E1	73.7	80.7
E2	69.6	71.8

there is a range of rank numbers where the error metrics are minimal. This sweet spot is where one finds optimal rank truncations that balances between having enough SVD modes included, while not oversaturating SHRED with noise and unpredictable small-scale turbulence.

2. Depth-dependent error error from data compression

To gain further insight into the effect of the compression, we isolate the error introduced by compression from that introduced by SHRED in two ways: First, we compare retained energy in the reconstructions over approximately one integral length scale in depth, with respect to the uncompressed data and the compressed data (computation details are reported in Sec. IV B). Second, we compute depth-dependent error metrics with the compressed data taken as ground truth and compare them to the error metrics with the full, uncompressed datasets as ground truth from Sec. IV B. Note that for the results presented here, we use the r values listed in Table II and represented by dashed lines in Fig. 13.

The retained energy in the reconstructed velocity field is reported in Table V, as a cumulative measure down to $z \approx L_\infty$, relative to the uncompressed and compressed fields. Measured against the compressed dataset, SHRED retains roughly 80% of the “kinetic energy” for data sets S1, S2 and E1, and 72% for dataset E2. When measured against the uncompressed data, the recovered cumulative kinetic energy is 2–7 percentage points lower. In short, SHRED underestimates the kinetic energy, with the main contribution coming from the reconstruction algorithm itself, not from the data compression step.

We next consider the depth-dependent metrics which compare errors introduced by SHRED and by the data compression, depicted in Fig. 14. In Figs. 14(a) and 14(b), we observe a results which basically reiterates the kinetic energy consideration above: The time-averaged u_{RMS} from the reconstructed fields is closer to the corresponding curve from the compressed data than that from the uncompressed data, but deviates more from both of these than they do from each other. In Figs. 14(c)–14(e), the general trend is, as expected, that the metrics for datasets S2 and E2 are shifted to lower errors and higher SSIM/PSNR, when measured against the compressed fields, rather than directly against the uncompressed data. Among these metrics, the NMSE and the SSIM stands out, in showcasing how much better performance we get when measuring against the compressed data (which is the data we actually train and validate on), rather than against the uncompressed data. Measured against the compressed data, the NMSE, the NMSE is reduced by roughly 0.04–0.07, equivalent to a 20%–40% reduction from the NMSE computed against the uncompressed data. A similar adjustment is seen for the SSIM, which is roughly 0.15–0.20 higher for the comparison with the compressed data. The difference can be attributed to the removal of small-scale structures by the compression, as seen in the field comparisons of Figs. 6 and 7, which is captured well by the NMSE and SSIM measures. The only measure that deviates from the trend is the PSNR for dataset E2, where the reconstruction performance is deemed to be slightly better when measured against the uncompressed rather than the compressed data.

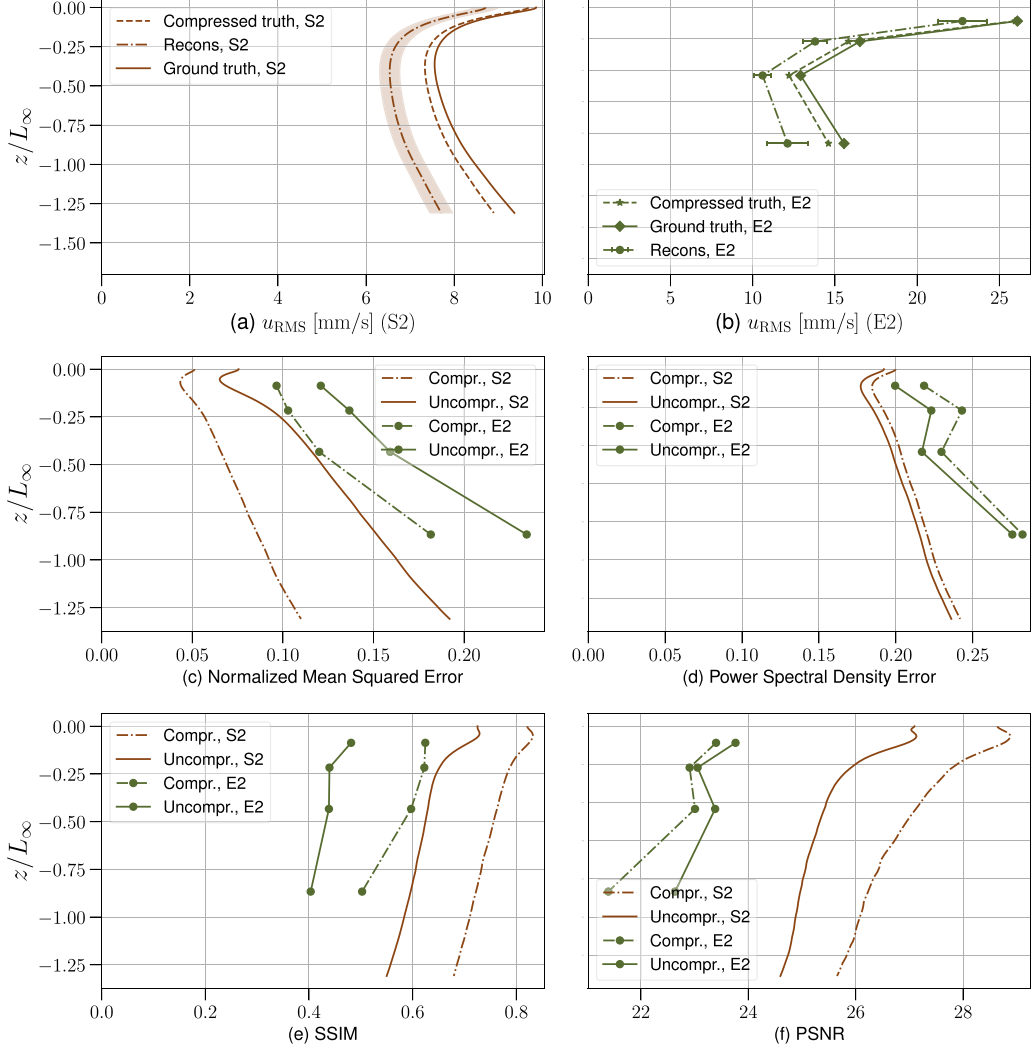


FIG. 14. Same comparison of error metrics as in Fig. 9, but comparing SHRED performance relative to the true value of the compressed data (dash-dotted curves) and to the uncompressed data (solid lines) for datasets S2 and E2. Panels (a) and (b) show reconstructed values of $\langle u_{RMS} \rangle$ for DNS and experimental datasets respectively, with true values for the compressed data added as dashed lines. The standard deviation of the ensembles are shown as shaded areas (a) and error bars (b). The bottom four panels show error metrics NMSE (c), PSDE (d), SSIM (e), and PSNR (f), all defined in Sec. III C.

APPENDIX B: INFLUENCE OF SENSOR COUNT ON SHRED PERFORMANCE

For the main analysis of SHRED, we used an ultra-sparse sensor scheme of three randomly placed surface sensors. SHRED has previously been shown remarkable reconstruction performance across different datasets with this sensor scheme using three sensors [28–30,35]. While this setup showcases the major strength of SHRED compared to other methods, e.g., POD-based regression, or even CNNs, the available sensor data might in some cases be dense, as in multiple sensor points or even high-resolution images. However, as SHRED is a nonlinear network, the effect of having more input data does not necessarily entail improved performance. In the paper where SHRED was first proposed, Williams *et al.* [28] analyzed reconstruction dependency on sensor count and

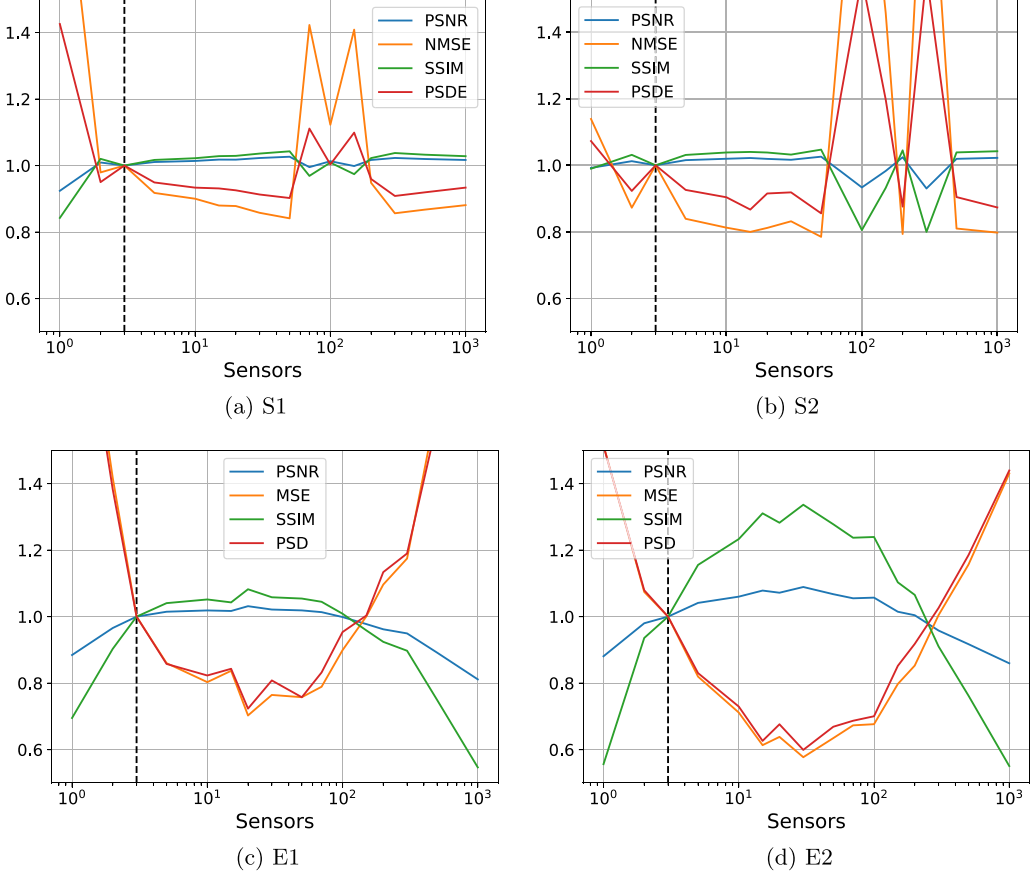


FIG. 15. Normalized error metrics averaged in depth for a range of sensor configurations. The dashed line marks the sensor configuration used throughout, which is three randomly placed points sensors at the surface. The error metrics are normalized based on the mean error at this configuration. Note that optimal SHRED performance is met if PSNR and SSIM are maximal, while the same is true when PSDE and NMSE are minimal.

placement, using up to 50 sensors. They found that, generally, the error decreased with increased number of sensors, and that random sensor placement had negligible impact on errors as compared to using a sophisticated QR-based sensor placement scheme. However, as the surface-to-subsurface sensing problem studied in the current paper differs from what has been considered in the previous SHRED application, and the limiting case with 50 sensors in [28] still uses a low number of sensors compared to available data points, we include a brief investigation of the reconstruction performance dependency on the sensor count here.

To investigate the impact the sensor count has on SHRED performance, we varied the number of sensors from 1 to 1000, and ran an ensemble of 20 trained SHRED models per sensor configuration. For each sensor configuration, we compute depth-averaged error metrics, in the same way as for the investigation of rank in Appendix A.1. The results are normalized by the results for the three-sensor configuration used throughout this paper. Figure 15 depicts the results for all four datasets. We first observe that using an order of magnitude more sensors would have had a positive impact on the results for all datasets. In particular, for the experimental datasets E1 and E2, increasing to 20–40 sensors appears to be a safe and beneficial choice, but going much beyond this, to, say, 100 sensors, leads to an increase in the error. For the DNS datasets, on the other hand, we observe that although

using 30 sensors reduces the reconstruction errors, a further increase in the number of sensors may lead to instabilities. This is evident in the abrupt jump in NMSE and PSDE for two of the tested sensor configurations for datasets S1 and S2, both outside the test regime reported in [28].

-
- [1] M. Brocchini and D. H. Peregrine, The dynamics of strong turbulence at free surfaces. Part 1. Description, *J. Fluid Mech.* **449**, 225 (2001).
 - [2] J. R. Aarnes, O. M. Babiker, A. Xuan, L. Shen, and S. Å. Ellingsen, Vortex structures under dimples and scars in turbulent free-surface flows, *J. Fluid Mech.* **1007**, A38 (2025).
 - [3] O. M. Babiker, I. Bjerkebæk, A. Xuan, L. Shen, and S. Å. Ellingsen, Vortex imprints on a free surface as proxy for surface divergence, *J. Fluid Mech.* **964**, R2 (2023).
 - [4] E. A. D’Asaro, Turbulence in the upper-ocean mixed layer, *Ann. Rev. Marine Sci.* **6**, 101 (2014).
 - [5] R. Wanninkhof, W. E. Asher, D. T. Ho, C. Sweeney, and W. R. McGillis, Advances in quantifying air-sea gas exchange and environmental forcing, *Annu. Rev. Mar. Sci.* **1**, 213 (2009).
 - [6] C. B. Brinkerhoff, C. J. Gleason, C. J. Zappa, P. A. Raymond, and M. E. Harlan, Remotely sensing river greenhouse gas exchange velocity using the SWOT satellite, *Global Biogeochem. Cycles* **36**, e2022GB007419 (2022).
 - [7] G. Dolcetti, B. Hortobágyi, M. Perks, S. J. Tait, and N. Dervilis, Using noncontact measurement of water surface dynamics to estimate river discharge, *Water Resour. Res.* **58**, e2022WR032829 (2022).
 - [8] N. B. Erichson, L. Mathelin, Z. Yao, S. L. Brunton, M. W. Mahoney, and J. N. Kutz, Shallow neural networks for fluid flow reconstruction with limited sensors, *Proc. R. Soc. A* **476**, 20200097 (2020).
 - [9] J. Zhang, J. Liu, and Z. Huang, Improved deep learning method for accurate flow field reconstruction from sparse data, *Ocean Eng.* **280**, 114902 (2023).
 - [10] P. Pant and A. B. Farimani, Deep learning for efficient reconstruction of high-resolution turbulent DNS data, *arXiv:2010.11348*.
 - [11] Z. Deng, C. He, Y. Liu, and K. C. Kim, Super-resolution reconstruction of turbulent velocity fields using a generative adversarial network-based artificial intelligence framework, *Phys. Fluids* **31**, 125111 (2019).
 - [12] S. Discetti and Y. Liu, Machine learning for flow field measurements: A perspective, *Meas. Sci. Technol.* **34**, 021001 (2022).
 - [13] Á. Moreno Soto, A. Güemes, and S. Discetti, Complete flow characterization from snapshot PIV, fast probes and physics-informed neural networks, *Comput. Methods Appl. Mech. Eng.* **419**, 116652 (2024).
 - [14] X. Jin, S. Laima, W.-L. Chen, and H. Li, Time-resolved reconstruction of flow field around a circular cylinder by recurrent neural networks based on non-time-resolved particle image velocimetry measurements, *Exp. Fluids* **61**, 114 (2020).
 - [15] Z. Deng, Y. Chen, Y. Liu, and K. C. Kim, Time-resolved turbulent velocity field reconstruction using a long short-term memory (LSTM)-based artificial intelligence framework, *Phys. Fluids* **31**, 075108 (2019).
 - [16] G. Borrelli, L. Guastoni, H. Eivazi, P. Schlatter, and R. Vinuesa, Predicting the temporal dynamics of turbulent channels through deep learning, *Int. J. Heat Fluid Flow* **96**, 109010 (2022).
 - [17] S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, Cambridge, 2019).
 - [18] P. A. Srinivasan, L. Guastoni, H. Azizpour, P. Schlatter, and R. Vinuesa, Predictions of turbulent shear flows using deep neural networks, *Phys. Rev. Fluids* **4**, 054603 (2019).
 - [19] S. L. Brunton, B. R. Noack, and P. Koumoutsakos, Machine learning for fluid mechanics, *Annu. Rev. Fluid Mech.* **52**, 477 (2020).
 - [20] A. Cuéllar, A. Güemes, A. Ianiro, Ó. Flores, R. Vinuesa, and S. Discetti, Three-dimensional generative adversarial networks for turbulent flow estimation from wall measurements, *J. Fluid Mech.* **991**, A1 (2024).
 - [21] M. Milano and P. Koumoutsakos, Neural network modeling for near wall turbulent flow, *J. Comput. Phys.* **182**, 1 (2002).

- [22] A. Xuan and L. Shen, Reconstruction of three-dimensional turbulent flow structures using surface measurements for free-surface flows based on a convolutional neural network, *J. Fluid Mech.* **959**, A34 (2023).
- [23] G. S. Hora, P. Gentine, M. Momen, and M. G. Giometto, Physics-informed data-driven reconstruction of turbulent wall-bounded flows from planar measurements, *Phys. Fluids* **36**, 115191 (2024).
- [24] K. Poulinakis, D. Drikakis, I. W. Kokkinakis, S. M. Spottswood, and T. Dbouk, LSTM reconstruction of turbulent pressure fluctuation signals, *Computation* **12**, 4 (2024).
- [25] S. Gakhar, J. R. Koseff, and N. T. Ouellette, Extracting free-surface expressions of underwater features, *Exp. Fluids* **63**, 138 (2022).
- [26] L. Guastoni, A. Güemes, A. Ianiro, S. Discetti, P. Schlatter, H. Azizpour, and R. Vinuesa, Convolutional-network models to predict wall-bounded turbulence from wall quantities, *J. Fluid Mech.* **928**, A27 (2021).
- [27] A. Güemes, S. Discetti, and A. Ianiro, Sensing the turbulent large-scale motions with their wall signature, *Phys. Fluids* **31**, 125112 (2019).
- [28] J. P. Williams, O. Zahn, and J. N. Kutz, Sensing with shallow recurrent decoder networks, *Proc. R. Soc. A* **480**, 20240054 (2024).
- [29] F. Faraji, M. Reza, and J. N. Kutz, Shallow recurrent decoder for reduced order modeling of $E \times B$ plasma dynamics, *Mach. Learn.: Sci. Technol.* **6**, 025024 (2025).
- [30] M. Tomasetto, J. P. Williams, F. Braghin, A. Manzoni, and J. N. Kutz, Reduced order modeling with shallow recurrent decoder networks, *Nat. Commun.* **16**, 10260 (2025).
- [31] F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics (Springer-Verlag, 1981), Vol. 898, pp. 366–381.
- [32] S. Riva, C. Introini, A. Cammi, and J. N. Kutz, Robust state estimation from partial out-core measurements with shallow recurrent decoder for nuclear reactors, *Prog. Nucl. Energy* **189**, 105928 (2025).
- [33] C. Introini, S. Riva, J. N. Kutz, and A. Cammi, From models to experiments: Shallow recurrent decoder networks on the DYNASTY experimental facility, [arXiv:2503.08907](https://arxiv.org/abs/2503.08907).
- [34] M. R. Ebers, J. P. Williams, K. M. Steele, and J. N. Kutz, Leveraging arbitrary mobile sensor trajectories with shallow recurrent decoder networks for full-state reconstruction, *IEEE Access* **12**, 97428 (2024).
- [35] M. L. Gao, J. P. Williams, and J. N. Kutz, Sparse identification of nonlinear dynamics and Koopman operators with shallow recurrent decoder networks, *Proc. Natl. Acad. Sci. USA* **123**, e2508144123 (2026).
- [36] J. R. Aarnes, O. M. Babiker, A. Xuan, L. Shen, and S. Å. Ellingsen, Replication data for: Vortex structures under dimples and scars in turbulent free-surface flows (2025), <https://doi.org/10.18710/XQ81WH>.
- [37] O. M. Babiker, J. R. Aarnes, A. Semati, A. Ferran, Y. H. Tee, R. J. Hearst, and S. Å. Ellingsen, Replication data for: Experimental investigation relating free-surface features to sub-surface turbulence (2026), <https://doi.org/10.18710/TSIZCQ>.
- [38] O. M. Babiker, J. R. Aarnes, A. Semati, A. Ferran, Y. H. Tee, R. J. Hearst, and S. Å. Ellingsen, Experimental investigation relating free-surface features to subsurface turbulence, *Phys. Rev. Fluids* **11**, 054802 (2026).
- [39] B. H. Brumley and G. H. Jirka, Near-surface turbulence in a grid-stirred tank, *J. Fluid Mech.* **183**, 235 (1987).
- [40] L. Chiapponi, S. Longo, and M. Tonelli, Experimental study on oscillating grid turbulence and free surface fluctuation, *Exp. Fluids* **53**, 1515 (2012).
- [41] H. Herlina and G. H. Jirka, Experiments on gas transfer at the air–water interface induced by oscillating grid turbulence, *J. Fluid Mech.* **594**, 183 (2008).
- [42] E. J. Hopfinger and J. A. Toly, Spatially decaying turbulence and its relation to mixing across density interfaces, *J. Fluid Mech.* **78**, 155 (1976).
- [43] T. Jamin, M. Berhanu, and E. Falcon, Experimental study of three-dimensional turbulence under a free surface, *Phys. Rev. Fluids* **10**, 034608 (2025).
- [44] T. Lacassagne, M. El-Hajem, F. Morge, S. Simoens, and J.-Y. Champagne, Study of gas liquid mass transfer in a grid stirred tank, *Oil Gas Sci. Tech.–Revue d’IFP Energies Nouvelles* **72**, 7 (2017).
- [45] S. P. McKenna and W. R. McGillis, The role of free-surface turbulence and surfactants in air–water gas transfer, *Int. J. Heat Mass Transf.* **47**, 539 (2004).

- [46] S. M. Thompson and J. S. Turner, Mixing across an interface due to turbulence generated by an oscillating grid, *J. Fluid Mech.* **67**, 349 (1975).
- [47] E. A. Variano and E. A. Cowen, Turbulent transport of a high-Schmidt-number scalar near an air–water interface, *J. Fluid Mech.* **731**, 259 (2013).
- [48] X. Guo and L. Shen, Interaction of a deformable free surface with statistically steady homogeneous turbulence, *J. Fluid Mech.* **658**, 33 (2010).
- [49] H. Herlina and J. G. Wissink, Direct numerical simulation of turbulent scalar transport across a flat surface, *J. Fluid Mech.* **744**, 217 (2014).
- [50] H. Herlina and J. G. Wissink, Simulation of air–water interfacial mass transfer driven by high-intensity isotropic turbulence, *J. Fluid Mech.* **860**, 419 (2019).
- [51] G. B. Folland, *Introduction to Partial Differential Equations*, 2nd ed. (Princeton University Press, Princeton, 1995).
- [52] J. N. Kutz, *Data-Driven Modeling & Scientific Computation: Methods for Complex Systems & Big Data* (Oxford University Press, 2013).
- [53] J. N. Kutz, M. Reza, F. Faraji, and A. Knoll, Shallow recurrent decoder for reduced order modeling of plasma dynamics, [arXiv:2405.11955](https://arxiv.org/abs/2405.11955).
- [54] https://github.com/jaarnes/fst_SHRED.
- [55] X. Guo and L. Shen, On the generation and maintenance of waves and turbulence in simulations of free-surface turbulence, *J. Comput. Phys.* **228**, 7313 (2009).
- [56] A. Xuan and L. Shen, A conservative scheme for simulation of free-surface turbulent and wave flows, *J. Comput. Phys.* **378**, 18 (2019).
- [57] A. Xuan and L. Shen, Analyses of wave-phase variation of Reynolds shear stress underneath surface wave using streamline coordinates, *J. Fluid Mech.* **931**, A32 (2022).
- [58] A. Semati, A. Shankaran, B. K. Smeltzer, E. Æsøy, R. J. Hearst, and S. Å. Ellingsen, Simultaneous free-surface profilometry and subsurface velocimetry with fringe projection and PIV, *Exp. Fluids* **67**, 120 (2026).
- [59] D. J. Ruth and F. Coletti, Structure and energy transfer in homogeneous turbulence below a free surface, *J. Fluid Mech.* **1001**, A46 (2024).
- [60] T. S. Lundgren, Linearly forced isotropic turbulence, Center for Turbulence Research Annual Research Briefs, 2023, <https://web.stanford.edu/group/ctr/ResBriefs03/lundgren.pdf>.
- [61] C. Rosales and C. Meneveau, Linear forcing in numerical simulations of isotropic turbulence: Physical space implementations and convergence properties, *Phys. Fluids* **17**, 095106 (2005).
- [62] G. Strang, *Linear Algebra and Learning from Data* (Cambridge Press, Wellesley, MA, 2019).
- [63] Z. Wang and A. C. Bovik, Mean squared error: Love it or leave it? A new look at signal fidelity measures, *IEEE Sign. Proc. Mag.* **26**, 98 (2009).
- [64] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Trans. on Image Process.* **13**, 600 (2004).
- [65] L. Shen, X. Zhang, D. K. P. Yue, and G. S. Triantafyllou, The surface layer for free-surface turbulent flows, *J. Fluid Mech.* **386**, 167 (1999).
- [66] S. Chen, T. Bao, P. Givi, C. Zheng, and X. Jia, Reconstructing turbulent flows using spatio-temporal physical dynamics, *ACM Trans. Intell. Syst. Technol.* **15**, 1 (2024).
- [67] X. Zhang, Y. Bao, M. L. Gao, and J. N. Kutz, Sendai: A hierarchical sparse-measurement, efficient data assimilation framework, [arXiv:2601.21664](https://arxiv.org/abs/2601.21664).