



OPEN

Identifying key differences between linear stochastic estimation and neural networks for fluid flow regressions

Taichi Nakamura¹, Kai Fukami^{1,2} & Koji Fukagata^{1✉}

Neural networks (NNs) and linear stochastic estimation (LSE) have widely been utilized as powerful tools for fluid-flow regressions. We investigate fundamental differences between them considering two canonical fluid-flow problems: (1) the estimation of high-order proper orthogonal decomposition coefficients from low-order their counterparts for a flow around a two-dimensional cylinder, and (2) the state estimation from wall characteristics in a turbulent channel flow. In the first problem, we compare the performance of LSE to that of a multi-layer perceptron (MLP). With the channel flow example, we capitalize on a convolutional neural network (CNN) as a nonlinear model which can handle high-dimensional fluid flows. For both cases, the nonlinear NNs outperform the linear methods thanks to nonlinear activation functions. We also perform error-curve analyses regarding the estimation error and the response of weights inside models. Our analysis visualizes the robustness against noisy perturbation on the error-curve domain while revealing the fundamental difference of the covered tools for fluid-flow regressions.

Recent boom of neural networks (NNs) has expanded into fluid mechanics^{1,2}. Although these have still been staying at the fundamental level, NNs have indeed shown their great potentials for data estimation^{3–5}, control⁶, reduced-order modeling⁷, and turbulence modeling⁸. Despite the recent enthusiastic trends on the use of NNs, we still have to rely largely on linear theories because their generalizability and transparency are, for the moment, superior to NN-based methods. This indicates that we may be able to obtain some clues from the relationships between NNs and linear theories to develop more advanced NN-based techniques for fluid flow analyses.

Motivated above, of particular interest here is an analogy between NNs and linear methods. Milano & Koumoutsakos⁹ used an autoencoder (AE) to perform low-dimensionalization for the Burgers equation and a turbulent channel flow. They also reported that the operation inside the multi-layer perceptron (MLP) based AE with linear activation functions is equivalent to that of proper orthogonal decomposition (POD)¹⁰. More recently, Murata et al.¹¹ investigated the analogy between a convolutional neural network (CNN)-based AE and POD. The strength of the nonlinear activation function used inside NNs was also demonstrated. From the perspective on state estimation, Nair & Goza¹² have recently compared the MLP, Gappy POD, and linear stochastic estimation (LSE) for the POD coefficient estimation from local sensor measurements of laminar wake of a flat plate.

To clarify the fundamental difference between linear methods and NNs, we here compare abilities of LSE, MLP, and CNN by considering two canonical fluid flow regression problems whose complexities are different with each other; 1. estimation of high-order POD coefficients from low-order counterparts of a flow around a cylinder¹³, and 2. state estimation from information on the wall in a turbulent channel flow¹⁴, as illustrated in Fig. 1. In particular, we seek key attributes for their estimations by focusing on influences of biases inside NNs, optimization methods, and robustness against noisy inputs.

Regression methods

Multi-layer perceptron. Let us first introduce a multi-layer perceptron (MLP)¹⁵. The MLP, which mimics the neurons in human's brain, has widely been utilized in physical sciences¹⁶. In the present paper, the MLP is used for low-dimensional fluid flow regressions, i.e., POD coefficient estimation, as an example of NN. The MLP is an aggregate of the minimum units called perceptrons. A linear superposition of input data q with weights W

¹Department of Mechanical Engineering, Keio University, Tokyo, Japan. ²Department of Mechanical and Aerospace Engineering, University of California, Los Angeles, USA. ✉email: fukagata@mech.keio.ac.jp

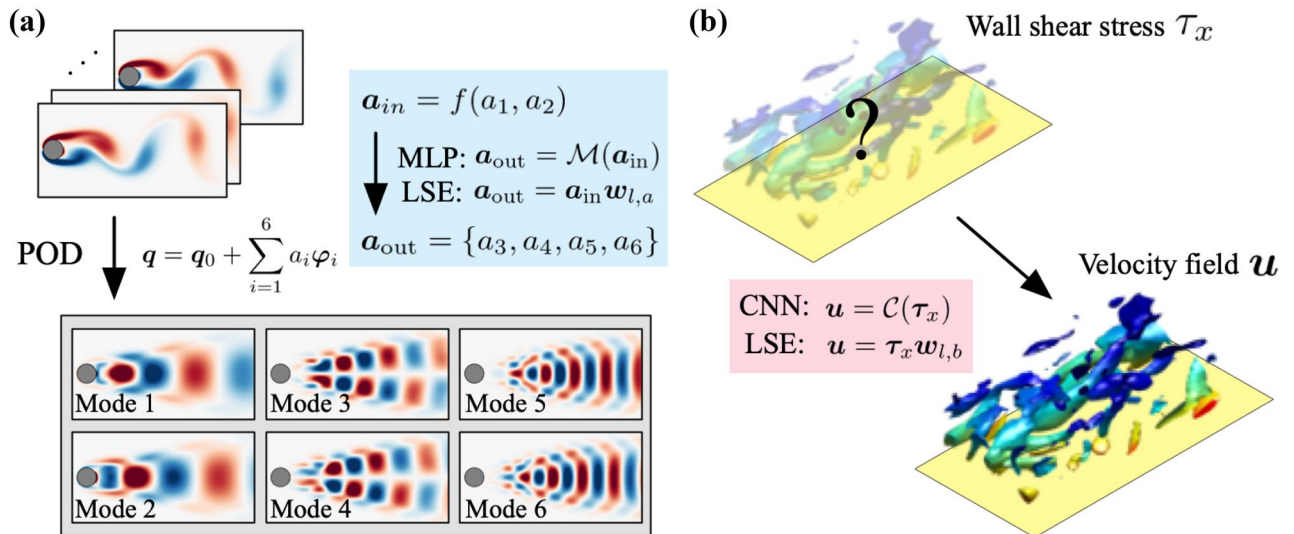


Figure 1. Covered fluid flow regressions in the present study. **(a)** POD coefficients estimation of a flow around a cylinder. **(b)** State estimation in a turbulent channel flow.

are passed inside a perceptron while biases **b** being added, and then a nonlinear activation function ϕ is applied such that

$$q_i^{(l)} = \phi \left(\sum_j W_{ij}^{(l)} q_j^{(l-1)} + b_j^{(l)} \right), \tag{1}$$

where l denotes a layer index. Weights among all edges W_{ij} are optimized with back-propagation¹⁸ so as to minimize a loss function E . The present MLP has hidden units of 4-8-16-8, while the number output nodes is 4 (3rd to 6th POD modes). The number of input nodes varies depending on considered cases, whose details will be provided in Sect. 3.1. In the simplest case, the present MLP \mathcal{M} attempts to output $\mathbf{a}_{out} = \{a_3, a_4, a_5, a_6\}$ from 2 inputs $\mathbf{a}_{in} = f(a_1, a_2)$, and the problem setting regarding weights \mathbf{w}_m (representing both \mathbf{W} and \mathbf{b}) inside the MLP can be represented as

$$\mathbf{w}_m = \operatorname{argmin}_{\mathbf{w}_m} \|\mathbf{a}_{out} - \mathcal{M}(\mathbf{a}_{in}; \mathbf{w}_m)\|_2. \tag{2}$$

We use the L_2 error norm as a loss function. Note that penalization terms, e.g., Lasso and Ridge penalties, are not considered in the present loss function because of the difficulty and cost in tuning the hyperparameter for regularization¹⁷.

Convolutional neural network. One of the issues associated with the MLP is that the number of edges inside it may explode due to its fully-connected structure when handling high-dimensional data such as fluid flows. To overcome this issue in dealing with fluid flow problems, a convolutional neural network (CNN)¹⁹ has widely been accepted as a good candidate^{3,20}. We capitalize on the combination of two- and three-dimensional CNNs for the state estimation task in the present study. The convolutional layer, which is a fundamental operation inside CNNs, extracts spatial features from input data using a filter operation,

$$q_{ijm}^{(l)} = \phi \left(b_m^{(l)} + \sum_{k=0}^{K-1} \sum_{p=0}^{H-1} \sum_{s=0}^{H-1} h_{pskm}^{(l)} q_{i+p-Cj+s-Ck}^{(l-1)} \right), \tag{3}$$

where $C = \text{floor}(H/2)$, K is the number of filters in a convolution layer, and $b_m^{(l)}$ is the bias. This filter operation achieves efficient data handling for two- or three-dimensional flow fields². For the three-dimensional CNN, the three-dimensional convolution is performed similarly to Eq. (3). In the present paper, the size H is set to 5. In the CNN, the output from the filtering operation is passed through an activation function ϕ , analogous to the operation inside an MLP (Eq. (1)). Filters h are optimized with back-propagation to minimize a loss function.

We use the combination of two- and three-dimensional CNNs to output a three-dimensional turbulent state \mathbf{u} from streamwise wall-shear stress $\tau_{x,wall}$ (details will be provided in Sect. 3.2). The optimization problem of weights \mathbf{w}_c (representing both \mathbf{h} and \mathbf{b}) for the CNN \mathcal{C} can be expressed as

$$\mathbf{w}_c = \operatorname{argmin}_{\mathbf{w}_c} \|\mathbf{u} - \mathcal{C}(\tau_{x,wall}; \mathbf{w}_c)\|_2. \tag{4}$$

Again, the aim of this study is the comparison between LSE and NN. To achieve a fair comparison, it is ideal to consider the same amount of weights for both LSE and NNs. To this end, we perform singular value decomposition (SVD) for the LSE weights \mathbf{w}_l to align the number of weights contained inside the CNNs and the LSE.

Layer	Data size	Layer	Data size
Input $\tau_{x,wall}$	(32, 32, 1)	14th 2D Conv.	(32, 32, 16)
1st 2D Conv.	(32, 32, 8)	15th 2D Conv.	(32, 32, 32)
2nd 2D Conv.	(32, 32, 8)	16th 2D Conv.	(32, 32, 32)
3rd 2D Conv.	(32, 32, 8)	Reshape	(32, 32, 32, 1)
4th 2D Conv.	(32, 32, 8)	1st 3D Conv.	(32, 32, 32, 8)
5th 2D Conv.	(32, 32, 8)	2nd 3D Conv.	(32, 32, 32, 8)
6th 2D Conv.	(32, 32, 16)	3rd 3D Conv.	(32, 32, 32, 16)
7th 2D Conv.	(32, 32, 16)	4th 3D Conv.	(32, 32, 32, 16)
8th 2D Conv.	(32, 32, 16)	5th 3D Conv.	(32, 32, 32, 8)
9th 2D Conv.	(32, 32, 16)	6th 3D Conv.	(32, 32, 32, 8)
10th 2D Conv.	(32, 32, 16)	7th 3D Conv.	(32, 32, 32, 8)
11th 2D Conv.	(32, 32, 16)	8th 3D Conv.	(32, 32, 32, 8)
12th 2D Conv.	(32, 32, 16)	9th 3D Conv.	(32, 32, 32, 3)
13th 2D Conv.	(32, 32, 16)	Output \mathbf{u}	(32, 32, 32, 3)

Table 1. Structure of 2D-3D CNN \mathcal{C} for turbulent channel flow example.

The details for the weight design for the LSE will be provided in the next section. The operation can be expressed as $\mathbf{w}_l = \mathbf{U}\mathbf{\Gamma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{n_{\text{input}} \times n_{\text{rank}}}$ and $\mathbf{V}^T \in \mathbb{R}^{n_{\text{rank}} \times n_{\text{output}}}$ are singular vectors, and $\mathbf{\Gamma} (\in \mathbb{R}^{n_{\text{rank}} \times n_{\text{rank}}})$ is a diagonal matrix whose diagonal elements are singular values. Since the rank of LSE weights in the present study is $n_{\text{rank}} = \text{rank}(\mathbf{w}_l) = 193$ according to our preliminary test, the present number of LSE weights based on SVD reduction $n_{\text{wLSE,SVD}}$ is 197,632 ($= n_{\text{input}} \times n_{\text{rank}} = 1024 \times 193$). Capitalizing on this SVD-based weight reduction, we are now able to obtain the target number of weights of approximately 197,000 to determine the CNN parameters. In this study, our CNN contains 196 608 weights with the parameters in Table 1.

Linear stochastic estimation. For comparison to the NNs, we use linear stochastic estimation (LSE)^{14,21}. In this study, we express target data $\mathbf{Q} \in \mathbb{R}^{n_{\text{data}} \times n_{\text{output}}}$ (output) as a linear map $\mathbf{w}_l \in \mathbb{R}^{n_{\text{input}} \times n_{\text{output}}}$ with respect to input data $\mathbf{P} \in \mathbb{R}^{n_{\text{data}} \times n_{\text{input}}}$ such that $\mathbf{Q} = \mathbf{P}\mathbf{w}_l$, where n_{data} represents the number of training snapshots, n_{input} represents the number of input attribute, and n_{output} represents the number of output attribute. Analogous to the optimization for NNs, the linear map \mathbf{w}_l can be obtained through a minimization such that

$$\mathbf{w}_l = \arg\min_{\mathbf{w}_l} \|\mathbf{Q} - \mathbf{w}_l\mathbf{P}\|_2 = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{Q}. \quad (5)$$

Note that penalization terms are not considered in the present loss function for the fair comparison to the NNs in terms of weight updating. We also emphasize that the LSE is analytically optimized solving Eq. (5) while the NNs are numerically optimized through back-propagation. Hence, we can also compare the NNs and the LSE regarding the difference of the optimization approaches. The optimized weights \mathbf{w}_l can then be applied to test data.

Results

Example 1: POD coefficient of two-dimensional cylinder wake at $Re_D = 100$. As presented in Fig. 1a, we first aim to estimate high-order POD coefficients $\mathbf{a}_{\text{out}} = \{a_3, a_4, a_5, a_6\}$ of a two-dimensional cylinder wake at $Re_D = 100$ from information of low-order counterparts $\mathbf{a}_{\text{in}} = f(a_1, a_2)$ such that $\mathbf{a}_{\text{out}} = \mathcal{F}_1(\mathbf{a}_{\text{in}})$, where \mathcal{F}_1 denotes a model for this purpose. The LSE and the MLP are used as the model \mathcal{F}_1 . Flow snapshots are generated using a two-dimensional direct numerical simulation (DNS). The governing equations are the incompressible Navier–Stokes equations,

$$\nabla \cdot \mathbf{u} = 0, \quad \partial_t \mathbf{u} + \nabla \cdot (\mathbf{u}\mathbf{u}) = -\nabla p + Re_D^{-1} \nabla^2 \mathbf{u}, \quad (6)$$

where \mathbf{u} and p denote the velocity vector and the pressure, respectively. All quantities are non-dimensionalized using the fluid density, the free-stream velocity, and the cylinder diameter. The size of the computational domain is $(L_x, L_y) = (25.6, 20.0)$, and the cylinder center is located at $(x, y) = (9, 0)$. The grid spacing and the time step are respectively $\Delta x = \Delta y = 0.025$ and $\Delta t = 2.5 \times 10^{-3}$, while imposing the no-slip boundary condition on the cylinder surface using an immersed boundary method²². The number of grid points used for the DNS is $(N_x, N_y) = (1024, 800)$. For the POD, the vorticity field ω around the cylinder is extracted as a domain of $8.2 \leq x \leq 17.8$ and $-2.4 \leq y \leq 2.4$ with $(N_x^*, N_y^*) = (384, 192)$.

We then take the POD for the collected snapshots to decompose the flow field \mathbf{q} as $\mathbf{q} = \mathbf{q}_0 + \sum_{i=1}^M a_i \varphi_i$, where φ denotes a POD basis, a is the POD coefficient, \mathbf{q}_0 is the temporal average of the flow field, and M represents the number of POD modes. For training the present MLP and LSE, we use 5000 snapshots. For comparison with LSE, we do not divide the training data for MLP into training and validation. We also consider additional 5000 snapshots for the assessment. We here compare the LSE with the linear MLP and the nonlinear MLP with ReLU activation function²³. The ReLU is known as a good candidate to prevent a vanishing gradient issue. We consider three patterns of input $\mathbf{a}_{\text{in}} = f(a_1, a_2)$ for the LSE and the linear MLP; the input $f(a_1, a_2)$ is $\mathbf{a}^{1\text{st}} = \{a_1, a_2\}$, or $\mathbf{a}^{2\text{nd}} = \{a_1, a_2, a_1 a_2, a_1^2, a_2^2\}$, or $\mathbf{a}^{3\text{rd}} = \{a_1, a_2, a_1 a_2, a_1^2, a_2^2, a_1^2 a_2, a_1 a_2^2, a_1^3, a_2^3\}$ while using only $\mathbf{a}^{1\text{st}}$ with the

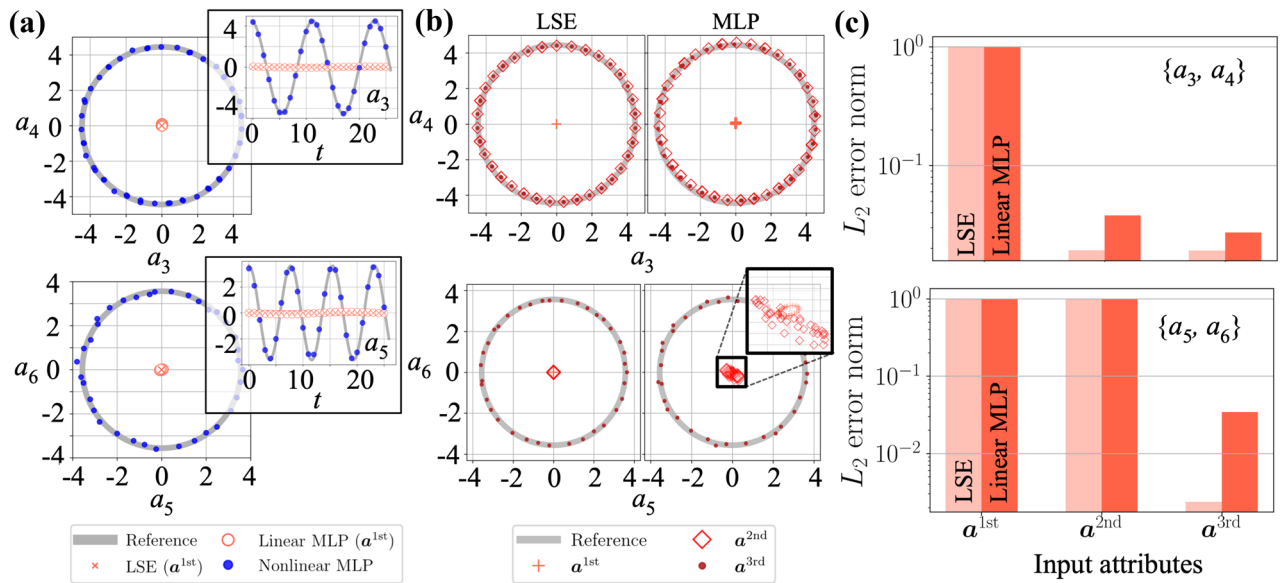


Figure 2. POD coefficients estimation from (a) 1st order coefficients $\mathbf{a}_{in} = \{a_1, a_2\}$ and (b) 2nd and 3rd order coefficients, and (c) the L_2 error norm $\varepsilon = \|\mathbf{a}_{out,ref} - \mathbf{a}_{out,est}\|_2 / \|\mathbf{a}_{out,ref}\|_2$.

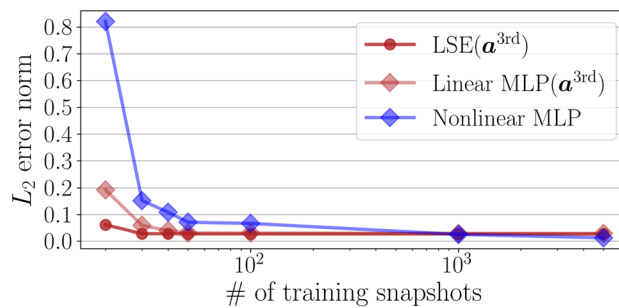


Figure 3. Dependence of the L_2 error norm for the POD coefficient estimation on the number of training snapshots.

nonlinear MLP. Since Loiseau et al.¹³ reported that the high-order coefficients \mathbf{a}_{out} can be represented using the quadratic expression of a_1 and a_2 due to its triadic interaction, this analysis enables us to check two viewpoints; 1. whether the nonlinear function inside MLP works to capture such nonlinear interactions, and 2. whether the linear models can also be utilized if a proper input including the essential combinations of nonlinear terms is given.

Let us demonstrate the estimation of $\mathbf{a}_{out} = \{a_3, a_4, a_5, a_6\}$ from only the information of first-order coefficients $\mathbf{a}^{1st} = \{a_1, a_2\}$ such that $\mathbf{a}_{out} = \mathcal{F}_1(\mathbf{a}^{1st})$, as shown in Fig. 2a. The nonlinear MLP shows its clear advantage against the LSE and the linear MLP for both coefficient maps. The L_2 error norm $\varepsilon = \|\mathbf{a}_{out,ref} - \mathbf{a}_{out,est}\|_2 / \|\mathbf{a}_{out,ref}\|_2$ for each case are 1.00 (LSE), 1.00 (linear MLP), and 0.0119 (nonlinear MLP), respectively. This suggests that the nonlinear activation function plays an important role in estimation. Noteworthy here, however, is that this nonlinearity can be recovered by giving a proper input, i.e., $\mathbf{a}_{in} = \{\mathbf{a}^{2nd}, \mathbf{a}^{3rd}\}$, even if we only use the linear methods, as presented in Figs. 2b,c. The reasonable estimation for $\{a_3, a_4\}$ can be achieved utilizing the input up to the 2nd order term \mathbf{a}^{2nd} , while that for $\{a_5, a_6\}$ requires the 3rd order term \mathbf{a}^{3rd} with both the LSE and the linear MLP. This trend is analogous to the observation by Loiseau et al.¹³, as introduced above. The LSE outperforms the linear MLP with the high-order coefficient inputs in this example, as shown in Fig. 2c; however, they will show a significant difference in terms of noise robustness as discussed later.

We then compare the LSE and the MLP in terms of the availability of training data. The dependence of the L_2 error on the number of training snapshots is examined in Fig. 3. Based on the results in Fig. 2, we choose the third-order coefficients $\mathbf{a}_{in} = \mathbf{a}^{3rd}$ for the linear models and the first-order coefficients for the nonlinear MLP, as the input for the models. The LSE shows its advantage over the MLP models when the number of training snapshots is limited. This is because the degree of freedom in the MLP is larger than that in the LSE. Note that the MLP will show a clear advantage against the LSE in terms of noise robustness, which reveals the fundamental difference between the linear MLP and the LSE.

We here consider the Gaussian white noise defined by the signal-to-noise ratio (SNR), $SNR = \sigma_{data}^2 / \sigma_{noise}^2$, where σ_{data}^2 and σ_{noise}^2 are the variances of input data and noise, respectively. The behaviors for noisy inputs are summarized in Fig. 4. As the linear models, we use the LSE and the linear MLP with $\mathbf{a}_{in} = \mathbf{a}^{3rd}$. For comparison,

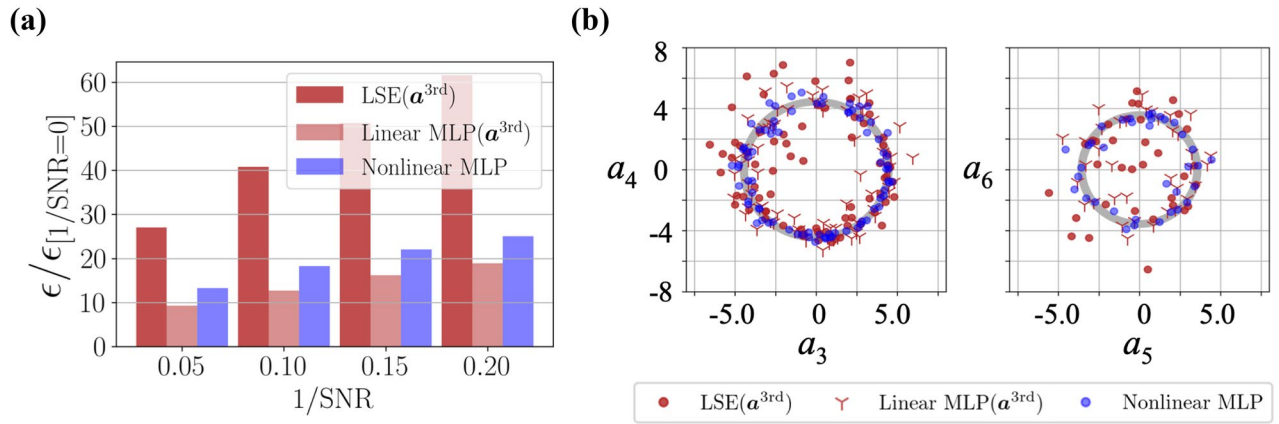


Figure 4. Robustness for noisy input of POD coefficients estimation. **(a)** Dependence of the increase ratio of the L_2 error norm $\epsilon/\epsilon_{[1/SNR=0]}$ on noise magnitude. **(b)** \mathbf{a}_{out} with $1/SNR = 0.1$ for the linear models with $\mathbf{a}_{in} = \mathbf{a}^{3rd}$ and the nonlinear MLP with $\mathbf{a}_{in} = \mathbf{a}^{1st}$.

we also monitor the nonlinear MLP with $\mathbf{a}_{in} = \mathbf{a}^{1st}$. The response of LSE is much more sensitive than that of the covered MLPs. This is caused by two considerable reasons: one is the influence of biases contained in the MLPs, as expressed in Eq. (1), while the other being the difference of optimization methods. Hereafter, we will seek which has the main contribution for the noise robustness.

For the investigation of the main contribution as mentioned above, we consider the LSE and three types of MLPs as follows:

1. LSE model: the same LSE model as that used above.
2. Linear MLP model with bias \mathcal{M}_1 : the same linear MLP as that used above
3. Linear MLP model without bias \mathcal{M}_2 : the biases are removed from the model \mathcal{M}_1 to investigate the influence on the bias.
4. Shallow linear MLP model without bias \mathcal{M}_3 : the MLP with a single hidden layer is prepared to align the number of weights with the LSE so that the difference of optimization methods can be assessed.

The dependence of the L_2 error norm on the noise magnitude of each model is summarized in Fig. 5a. There is almost no difference among the covered MLP models. This suggests that the bias and the number of layers do not contribute to the noise robustness so much. The other observation here is that the shallow linear MLP is still more robust than the LSE, even though the model structures are identical with each other. To examine this point, we visualize the weights inside the LSE and the shallow linear MLP in Fig. 5b. The weights for the second-order term input ($a_1 a_2, a_1^2, a_2^2$) are optimized to the same values with each other, while that for the first-order input term exhibits a significant difference. This is caused by the difference of the optimization methods. Moreover, this point is examined by visualizing an error surface for the input of a_1 and a_1^3 with the output of a_5 , as shown in Fig. 5d. The reason for the choice of this input-output combination is that the output a_5 is one of the most sensitive components to the noise for the LSE, as presented in Fig. 5c. The optimized solutions are different to each other, which is likely caused by the difference in optimization methods. What is notable here is that the noise addition drastically changes the error-surface shape of the LSE, while that of the MLP changes only slightly. This difference can be quantified using the mean structural similarity index measure (MSSIM)²⁴. We apply this to the elevation of each error surface, i.e., without $E_r \in \mathbb{R}^{M \times N}$ and with noise $E_n \in \mathbb{R}^{M \times N}$, where M, N are the number of samples on the error surface for each weight axis. MSSIM is mathematically expressed as

$$MSSIM(E_r, E_n) = \frac{1}{M'N'} \sum_{i=1}^{M'} \sum_{j=1}^{N'} SSIM(e_{r,ij}, e_{n,ij}), \quad SSIM(e_r, e_n) = \frac{(2\mu_r\mu_n + C_1)(2\sigma_m + C_2)}{(\mu_r^2 + \mu_n^2 + C_1)(\sigma_r^2 + \sigma_n^2 + C_2)}. \tag{7}$$

This measurement can assess a similarity between two images $E_r \in \mathbb{R}^{M \times N}$ and $E_n \in \mathbb{R}^{M \times N}$ by considering their mean μ and standard deviation σ . To obtain MSSIM, the SSIM in a small window of two images $e_r \in \mathbb{R}^{m \times n}$ and $e_n \in \mathbb{R}^{m \times n}$, where $M' = M - m + 1$ and $N' = N - n + 1$, is computed and its average is taken over the image. As the constant values C_1 and C_2 in Eq. (7), we set $\{C_1, C_2\} = \{0.16, 1.44\}$ following Wang et al.²⁴. As presented in Fig. 5d, the MSSIM of the shallow linear MLP reports 0.879 while that of LSE is 0.428, which indicates that the deformation of the error surface is substantially larger in the LSE. Due to this large deformation of the error surface, the optimum point of the LSE is pushed up vertically in the error space of Fig. 5d. This indicates that the weights obtained by the LSE in an analytical manner guarantee the global optimal solution over the training data; however, this solution may not be optimal from the viewpoint of noise robustness. On the other hand, the MLP provides a noise-robust solution, although it is not the exact global optimum over the training data since the MLP weights are a numerical solution obtained through back-propagation.

Considering the characteristics of LSE which fits training data, we then add noisy data of $SNR = 0.05$ to training data for both LSE and the shallow linear MLP (without bias), as summarized in Fig. 6. As shown in Fig. 6a,

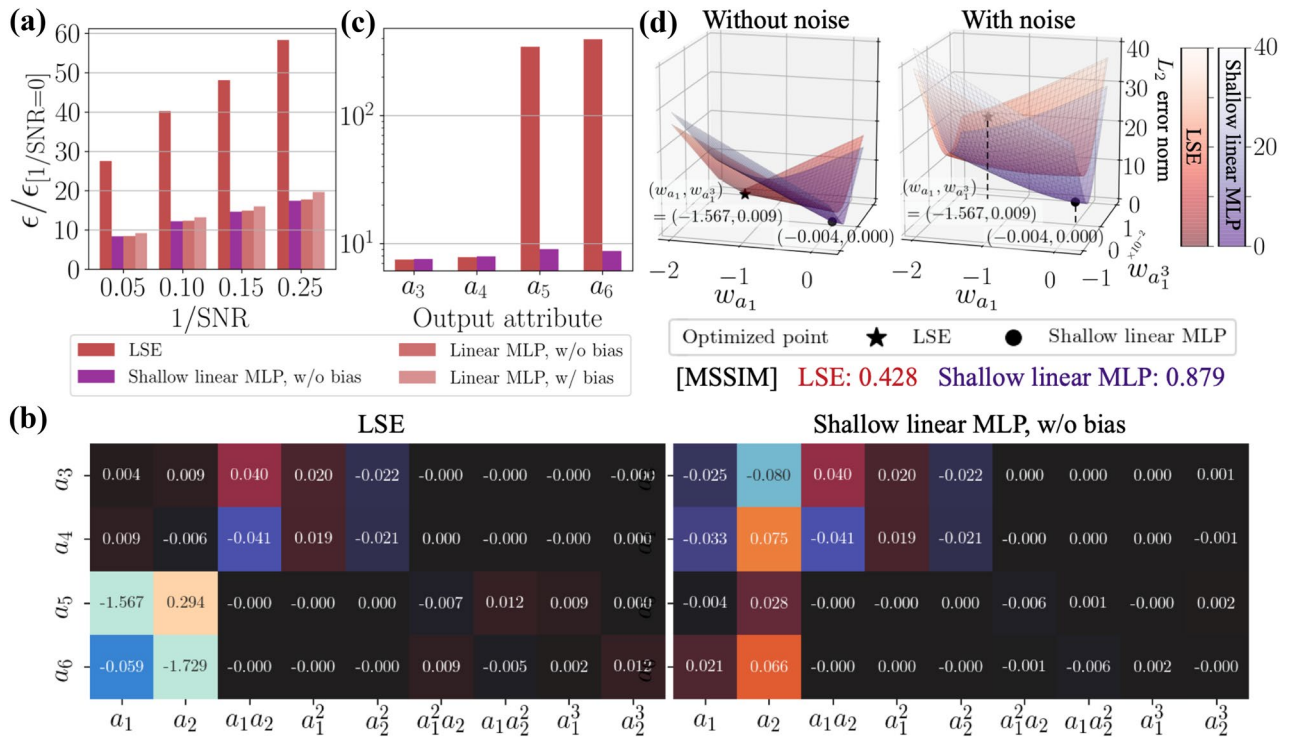


Figure 5. Comparison of the LSE and the linear MLP focusing on the bias and the optimization method. (a) Increasing ratio $\epsilon/\epsilon_{[1/SNR=0]}$ of the L_2 error norm ϵ from the original error $\epsilon_{[1/SNR=0]}$ without the noisy input. (b) Weight values of the LSE and the shallow linear MLP. (c) Dependence of the increasing ratio $\epsilon/\epsilon_{[1/SNR=0]}$ of the L_2 error norm at $1/SNR = 0.05$ on the output POD coefficients. (d) Visualization of the error surface around the optimized point.

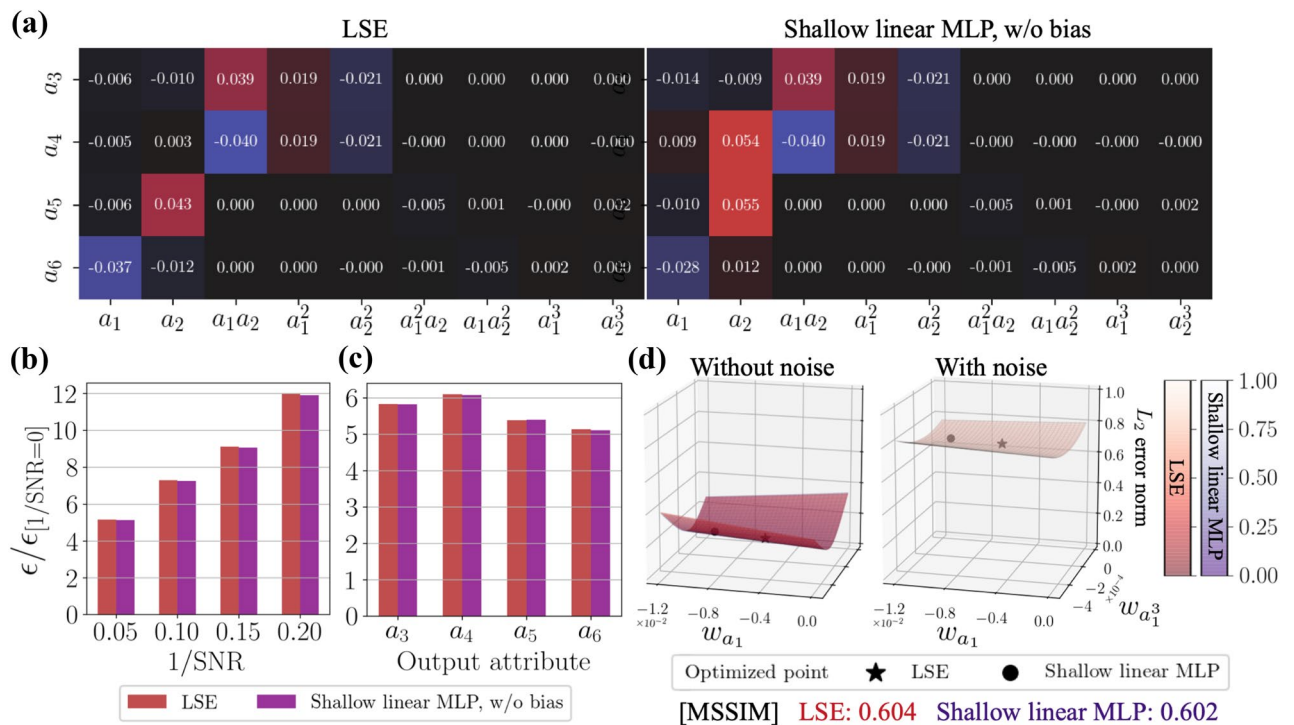


Figure 6. Comparison of the LSE and the linear MLP trained with the noisy training data. (a) Weight values of the LSE and the shallow linear MLP. (b) Increasing ratio $\epsilon/\epsilon_{[1/SNR=0]}$ of the L_2 error norm ϵ from the original error $\epsilon_{[1/SNR=0]}$ without the noisy input. (c) Dependence of the increasing ratio $\epsilon/\epsilon_{[1/SNR=0]}$ of the L_2 error norm at $1/SNR = 0.05$ on the output POD coefficients. (d) Visualization of the error surface around the optimized point.

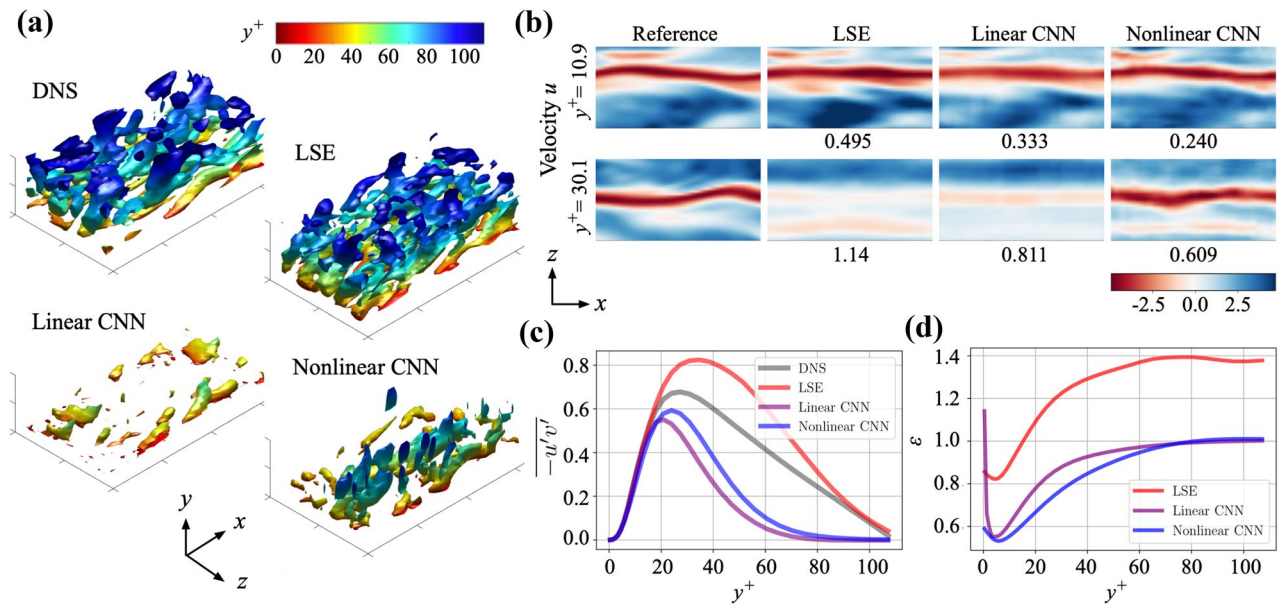


Figure 7. Estimation of turbulent channel flow from streamwise wall-shear stress. **(a)** Isosurfaces of the Q criterion ($Q^+ = -0.005$). **(b)** $x - z$ sectional velocities at $y^+ = 10.9$ and 30.1 . The values underneath the contours report the L_2 error norm for each velocity attribute. **(c)** Reynolds shear stress $-\overline{u'v'}$. **(d)** Dependence of the ensemble L_2 error norm over three velocity components on the y position.

there is almost no difference in terms of weight values despite that there was in Fig. 5. The similar observations can also be found for several analyses in Figs. 6b–d. These suggest that the LSE can also obtain robustness by adding noise to the training data. Note that the increasing ratio $\varepsilon/\varepsilon_{[1/SNR=0]} = 5$ in Fig. 6b is not quite large because the original error $\varepsilon_{[1/SNR=0]}$ is small ($\varepsilon_{[1/SNR=0]} = 4.49 \times 10^{-2}$ for the LSE and $\varepsilon_{[1/SNR=0]} = 4.52 \times 10^{-2}$ for the shallow linear MLP).

Example 2: velocity field in a minimal turbulent channel flow at $Re_\tau = 110$. We then perform the comparison of NNs and LSE for a more complex problem. Let us consider the estimation of a velocity field \mathbf{u} in a minimal turbulent channel flow at $Re_\tau = 110$ from the streamwise wall shear stress input $\boldsymbol{\tau}_x$, as illustrated in Fig. 1b, such that $\mathbf{u} = \mathcal{F}_2(\boldsymbol{\tau}_x)$, where \mathcal{F}_2 denotes a model for example 2. As already mentioned, we use LSE and CNN as the model \mathcal{F}_2 . The training data is generated using a three-dimensional DNS which numerically solves the incompressible Navier–Stokes equations,

$$\nabla \cdot \mathbf{u} = 0, \quad \partial_t \mathbf{u} + \nabla \cdot (\mathbf{u}\mathbf{u}) = -\nabla p + Re_\tau^{-1} \nabla^2 \mathbf{u}, \tag{8}$$

where \mathbf{u} and p represents the velocity vector and pressure, respectively^{25,26}. The quantities are non-dimensionalized with the channel half-width δ and the friction velocity u_τ . The computational domain is $(L_x, L_y, L_z) = (\pi\delta, 2\delta, 0.5\pi\delta)$ with the number of grid points of $(N_x, N_y, N_z) = (32, 64, 32)$. The grid is arranged uniformly in the x and the z directions, while nonuniformly in the y direction. The time step is $\Delta t^+ = 0.0385$, where the subscript $+$ denotes the wall units. We use 10,000 snapshots to train the models. For equivalent comparison with LSE, we do not divide the training data into training and validation. We also prepare additional 2700 snapshots for the assessment.

The channel flow fields estimated by the models are assessed in Fig. 7a. Note again that the number of weights inside the CNN and the LSE is almost the same with each other as explained in Sect. 2.2. We here also consider the linear CNN (i.e., the same CNN structure, but with the linear activation function) to examine whether the similar observation to the linear MLP of the cylinder example can be found or not even for a CNN whose filter operation is different from a fully-connected MLP. The estimated fields are visualized using the second invariant of the velocity gradient tensor Q^+ of -0.005 . The field reconstructed by the LSE shows the similar behavior to the reference DNS data qualitatively, e.g., the amount of vortical structure; however, it should be emphasized that the LSE field merely provides turbulent-like structure, which is different from that of the DNS.

To investigate this point, we also visualize the $x - z$ sectional streamwise velocity distributions in Fig. 7b. Notably, the nonlinear CNN outperforms the LSE and the linear CNN, which is not intuitive from the observation with the Q isosurfaces. Especially, the advantage of the nonlinear method can be clearly found at $y^+ = 30.1$. We also present the time-averaged Reynolds shear stress $-\overline{u'v'}$ in Fig. 7c. The LSE curve looks to be in reasonable agreement in its shape (although overestimated), despite its high-error level as stated above. With the observation for the reasonable reconstruction of LSE in Fig. 7a, it implies that the flow field estimated by the LSE model is similar to the DNS data in a time-ensemble sense, although it does not match for each instantaneous field.

We also investigate the dependence of estimation on the y position in Fig. 7d. Analogous to previous studies^{14,27}, it is hard to estimate the velocity field in the region away from the wall because of the lack of

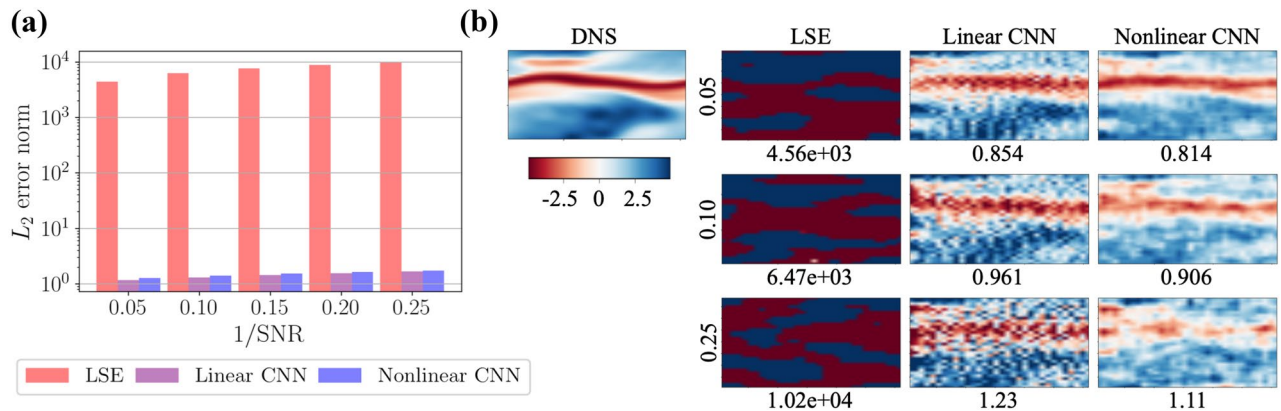


Figure 8. Robustness for noisy input in the turbulent channel flow example. (a) $\varepsilon/\varepsilon_{[1/\text{SNR}=0]}$ on noise magnitude. (b) Contours of estimated streamwise velocity fluctuation u' at $y^+ = 10.9$. The values underneath the contours indicate the ensemble L_2 error norm over three velocity components. The numbers on the side of the contour represent $1/\text{SNR}$.

correlation between the wall shear stress τ_x and the velocity field \mathbf{u} away from the wall. Within the range that can be estimated, the nonlinear CNN presents the better estimation than the linear methods in terms of both the L_2 error and the statistics.

We further assess the difference among the linear and nonlinear methods focusing on noise robustness, as summarized in Fig. 8. Analogous to the noise investigation for the POD coefficient estimation, the noisy input for the wall shear stress is designed with the SNR. The L_2 error of the LSE explodes rapidly with the noise addition, which is the same behavior as that in Fig. 4a. In contrast, the CNNs are still able to reconstruct the large scale structure despite that the LSE cannot even with $1/\text{SNR} = 0.05$. These findings also make us to suspect the role of optimization methods.

Let us then discuss the contribution of optimization methods for noise robustness. We here skip the influence on bias uses since there is no significant effects according to our preliminary test, which is akin to the trend with the cylinder example. For the particular demonstration here, we arrange a shallow linear CNN, in which there are n_{output} filters and each filter has a shape of $(n_{\text{input}}, 1)$. Note that this shallow CNN is distinct against the CNNs used above, which were based on the original SVD-based weight reduction expressed in Sect. 2.2. The use of the shallow linear CNN enables us to observe the weight sensitivity against noisy inputs while comparing to the LSE directly, thanks to its filter shape.

The error surfaces of the LSE and the shallow linear CNN are visualized in Figs. 9a–c. Note that we here choose six points for the visualization. The error used in the error surface is arranged by the streamwise velocity u at an arbitrary point in the $x-z$ cross section at $y^+ = 15.4$. In the case of the shallow linear CNN, the noise has little influence on the error surface. On the other hand, the error surfaces of the LSE drastically change their shape with the existence of noise. These trends can also be found with the MSSIM. It implies that the CNN can also obtain noise robustness thanks to the gradient method, while it is hard to obtain with the LSE, which is analogous to the observation with the POD coefficient estimation.

Conclusions

Fundamental differences between neural networks (NNs) and linear methods were investigated considering canonical fluid flow regression problems: 1. the estimation of POD coefficients of a flow around a cylinder, and 2. the state estimation from wall measurements in a turbulent channel flow. We compared linear stochastic estimation (LSE) with multi-layer perceptron (MLP) and convolutional neural network (CNN). For both regression problems, efficacy of nonlinear function can be observed. We also found that a linear model could surrogate a nonlinear model by giving an appropriate combination of inputs under the consideration of nonlinear relationship. This enables us to expect that the combination of nonlinear activation functions and proper inputs can further enhance the prediction capability of the model, which is similar to the observation in several previous studies^{28–30}. In addition, the linear NNs were more robust against noise than the LSE, and the reason for this was revealed by visualizing the error surface. The error surface told us that the difference in optimization methods has a significant contribution to the noise robustness.

Although we observed the strength of nonlinear NNs from several perspectives, we should note that the learning process of NNs can also be unstable depending on a problem setting since it is founded on a gradient method. This implies that we may not reach a reasonable valley of a weight surface, especially when the problem has a multimodality and the availability of training data is limited³¹. In this sense, the LSE can provide us a stable solution in a theoretical manner. Hence, we may be able to unify these characteristics for further improvement of NN learning pipelines, e.g., transfer learning^{32,33}, so that a learning process of NN can be started from a reasonable solution while achieving a proper noise robustness. Furthermore, it should also be emphasized that we can introduce a loss function associated with a priori knowledge from physics since both NNs and LSE are based on a minimization manner in terms of weights. Inserting a physical loss function may be one of the considerable paths towards practical applications of both methods^{34–36}.

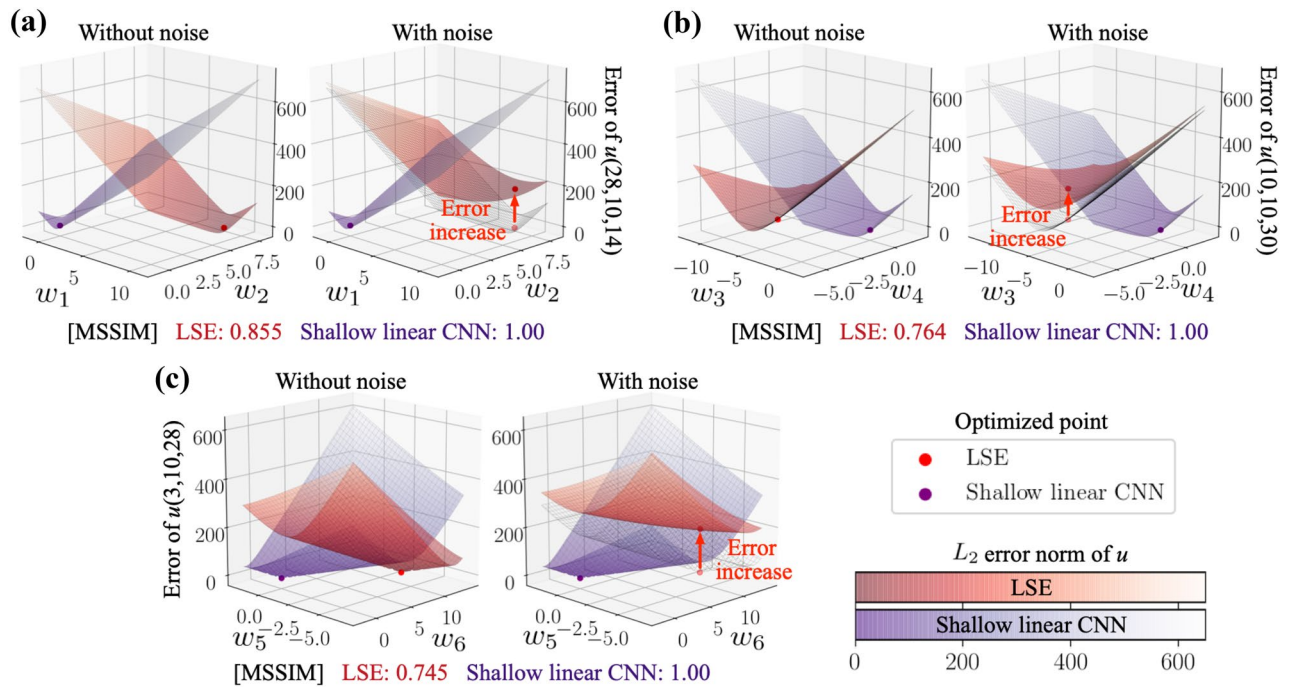


Figure 9. Error-surface analysis in the turbulent channel flow example. The axis x and y are arbitrary weights picked up from each model.

Received: 20 May 2021; Accepted: 18 February 2022
 Published online: 08 March 2022

References

- Brunton, S. L., Noack, B. R. & Koumoutsakos, P. Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.* **52**, 477–508 (2020).
- Fukami, K., Fukagata, K. & Taira, K. Assessment of supervised machine learning for fluid flows. *Theor. Comp. Fluid Dyn.* **34**, 497–519 (2020).
- Fukami, K., Fukagata, K. & Taira, K. Super-resolution reconstruction of turbulent flows with machine learning. *J. Fluid Mech.* **870**, 106–120 (2019).
- Fukami, K., Fukagata, K. & Taira, K. Machine-learning-based spatio-temporal super resolution reconstruction of turbulent flows. *J. Fluid Mech.* **909**, A9. <https://doi.org/10.1017/jfm.2020.948> (2021).
- Fukami, K., Maulik, K., Ramachandra, N., Fukagata, K. & Taira, K. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nat. Mach. Intell.* **3**, 945–951 (2021).
- Brenner, M. P., Eldredge, J. D. & Freund, J. B. Perspective on machine learning for advancing fluid mechanics. *Phys. Rev. Fluids* **4**, 100501. <https://doi.org/10.1103/PhysRevFluids.4.100501> (2019).
- Maulik, R., Fukami, K., Ramachandra, N., Fukagata, K. & Taira, K. Probabilistic neural networks for fluid flow surrogate modeling and data recovery. *Phys. Rev. Fluids* **5**, 104401. <https://doi.org/10.1103/PhysRevFluids.5.104401> (2020).
- Duraisamy, K., Iaccarino, G. & Xiao, H. Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.* **51**, 357–377 (2019).
- Milano, M. & Koumoutsakos, P. Neural network modeling for near wall turbulent flow. *J. Comput. Phys.* **182**, 1–26 (2002).
- Lumley, J. L. The structure of inhomogeneous turbulent flows. In Yaglom, A. M. & Tatarski, V. I. (eds.) *Atmospheric turbulence and radio wave propagation* (Nauka, 1967).
- Murata, T., Fukami, K. & Fukagata, K. Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *J. Fluid Mech.* **882**, A13. <https://doi.org/10.1017/jfm.2019.822> (2020).
- Nair, N. J. & Goza, A. Leveraging reduced-order models for state estimation using deep learning. *J. Fluid Mech.* **897**. <https://doi.org/10.1017/jfm.2020.409> (2020).
- Loiseau, J.-C., Brunton, S. L. & Noack, B. R. *From the POD-Galerkin method to sparse manifold models*. <https://doi.org/10.13140/RG.2.2.27965.31201> (2018).
- Suzuki, T. & Hasegawa, Y. Estimation of turbulent channel flow at $Re_\tau = 100$ based on the wall measurement using a simple sequential approach. *J. Fluid Mech.* **830**, 760–796 (2006).
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagation errors. *Nature* **322**, 533–536 (1986).
- Domingos, P. A few useful things to know about machine learning. *Commun. ACM* **55**, 78–87 (2012).
- Nakamura, T., & Fukagata, K. Robust training approach of neural networks for fluid flow state estimations. Preprint at [arXiv:2112.02751](https://arxiv.org/abs/2112.02751) (2021).
- Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. Preprint at [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
- Morimoto, M., Fukami, K., Zhang, K., Nair, A. G. & Fukagata, K. Convolutional neural networks for fluid flow analysis: Toward effective metamodeling and low-dimensionalization. *Theor. Comp. Fluid Dyn.* **35**, 633–658 (2021).
- Adrian, R. J. & Moin, P. Stochastic estimation of organized turbulent structure: homogeneous shear flow. *J. Fluid Mech.* **190**. <https://doi.org/10.1017/S0022112088001442> (1988).
- Kor, H., Badri Ghomizad, M. & Fukagata, K. A unified interpolation stencil for ghost-cell immersed boundary method for flow around complex geometries. *J. Fluid Sci. Technol.* **12**, JFST0011. <https://doi.org/10.1299/jfst.2017jfst0011> (2017).
- Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. *Proc. Int. Conf. Mach. Learn.* 807–814. <https://doi.org/10.5555/3104322.3104425> (2010).

24. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* **13**, 600–612 (2004).
25. Fukagata, K., Kasagi, N. & Koumoutsakos, P. A theoretical prediction of friction drag reduction in turbulent flow by superhydrophobic surfaces. *Phys. Fluids* **18**, 051703 (2006).
26. Nakamura, T., Fukami, K., Hasegawa, K., Nabaie, Y. & Fukagata, K. Convolutional neural network and long short-term memory based reduced order surrogate for minimal turbulent channel flow. *Phys. Fluids* **33**, 025116 (2021).
27. Chevalier, M., Hoepffner, J., Bewley, T. R. & Henningson, D. S. State estimation in wall-bounded flow systems. part 2. turbulent flows. *J. Fluid Mech.* **552**, 167–187. <https://doi.org/10.1017/S0022112005008578> (2006).
28. Gamahara, M. & Hattori, Y. Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids* **2**(5), 054604 (2017).
29. Kim, J. & Lee, C. Prediction of turbulent heat transfer using convolutional neural networks. *J. Fluid Mech.* **882**, A18 (2020).
30. Park, J. & Choi, H. Toward neural-network-based large eddy simulation: Application to turbulent channel flow. *J. Fluid Mech.* **914**, A16 (2021).
31. Morimoto, M., Fukami, K., Maulik, R., Vinuesa, R., & Fukagata, K. Assessments of model-form uncertainty using Gaussian stochastic weight averaging for fluid-flow regression. Preprint at [arXiv:2109.08248](https://arxiv.org/abs/2109.08248) (2021).
32. Morimoto, M., Fukami, K., Zhang, K. & Fukagata, K. Generalization techniques of neural networks for fluid flow estimation. *Neural Comput. Appl.* <https://doi.org/10.1007/s00521-021-06633-z> (2021).
33. Guastoni, L. *et al.* Convolutional-network models to predict wall-bounded turbulence from wall quantities. *J. Fluid Mech.* **928**, A27. <https://doi.org/10.1017/jfm.2021.812> (2021).
34. Lee, S. & You, D. Data-driven prediction of unsteady flow fields over a circular cylinder using deep learning. *J. Fluid Mech.* **879**, 217–254. <https://doi.org/10.1017/jfm.2019.700> (2019).
35. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
36. Du, Y. & Zaki, T. A. Evolutional deep neural network. *Phys. Rev. E* **104**, 045303 (2021).

Acknowledgements

This work was supported through JSPS KAKENHI Grant Number 18H03758 and 21H05007 by Japan Society for the Promotion of Science.

Author contributions

T.N and Ka.F designed research; T.N. performed research and analyzed data; Ka.F. and Ko.F. supervised. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ko.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022