AIP Advances

Prospects of federated machine learning in fluid dynamics

Cite as: AIP Advances **12**, 095212 (2022); https://doi.org/10.1063/5.0104344
Submitted: 19 June 2022 • Accepted: 17 August 2022 • Published Online: 13 September 2022

Omer San, Suraj Pawar and Adil Rasheed

AIP Advances ARTICLE scitation.org/journal/adv

Prospects of federated machine learning in fluid dynamics

Cite as: AIP Advances 12, 095212 (2022); doi: 10.1063/5.0104344
Submitted: 19 June 2022 • Accepted: 17 August 2022 •
Published Online: 13 September 2022

View Online Export Citation CrossMark

Omer San,[1,a)] Suraj Pawar,[1] and Adil Rasheed[2]

AFFILIATIONS

[1] School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, Oklahoma 74078, USA
[2] Department of Engineering Cybernetics, Norwegian University of Science and Technology, N-7465 Trondheim, Norway

[a)] Author to whom correspondence should be addressed: osan@okstate.edu

ABSTRACT

Physics-based models have been mainstream in fluid dynamics for developing predictive models. In recent years, machine learning has offered a renaissance to the fluid community due to the rapid developments in data science, processing units, neural network based technologies, and sensor adaptations. So far in many applications in fluid dynamics, machine learning approaches have been mostly focused on a standard process that requires centralizing the training data on a designated machine or in a data center. In this article, we present a federated machine learning approach that enables localized clients to collaboratively learn an aggregated and shared predictive model while keeping all the training data on each edge device. We demonstrate the feasibility and prospects of such a decentralized learning approach with an effort to forge a deep learning surrogate model for reconstructing spatiotemporal fields. Our results indicate that federated machine learning might be a viable tool for designing highly accurate predictive decentralized digital twins relevant to fluid dynamics.

INTRODUCTION

In many complex systems involving fluid flows, computing a physics-based model might be prohibitive, especially when our simulations are compatible with the timescales of natural phenomena. Consequently, there is an ever-growing interest in generating surrogate or reduced order models.[1] It has also been envisioned that a digital twin capable of accurately representing the physical system could offer a better value proposition to specific applications and stakeholders.[2] The role of this digital twin might be to provide descriptive, diagnostic, predictive, or prescriptive guidelines for a better-informed decision. The *market pull* created by digital twin-like technologies coupled with the *technology push* provided by significant advances in machine learning (ML) and artificial intelligence (AI), advanced and cost-effective sensor technologies, readily available computational resources, and opensource ML libraries have accelerated ML penetration in domain sciences like never before. The last decade has seen an exponential growth of data-driven modeling technologies (e.g., deep neural networks) that might be key enablers for improving the modeling accuracy of geophysical fluid systems.[3] A recent workshop held by the NASA Advanced Information Systems Technology Program and Earth Science Information Partners on ML adoption[4] identified the following guidelines, among many others, in this area:

- Cutting edge ML algorithms and techniques need to be available, packaged in some way and well understood so as to be usable.
- Computer security implementations are outdated and uncooperative with science investigations. Research in making computational resources secure and yet easily usable would be valuable.

One of the fluid flow problems that ML and AI can positively impact is weather forecasting. Big data will be the key to making the digital twins of the natural environments a reality. In addition to the data from forecasting models and dedicated weather stations, it can be expected that there will be an unprecedented penetration of smart devices (e.g., smart weather stations, smartphones, and smartwatches) and contributions from crowdsourcing. For example, by 2025, there will be more than 7×10^9 smartphones worldwide. This number is much more significant than the paltry (over 10 000) official meteorological stations around the world.[5] While analyzing and utilizing data only from a few edge devices might not yield accurate predictions, processing data from many smart and

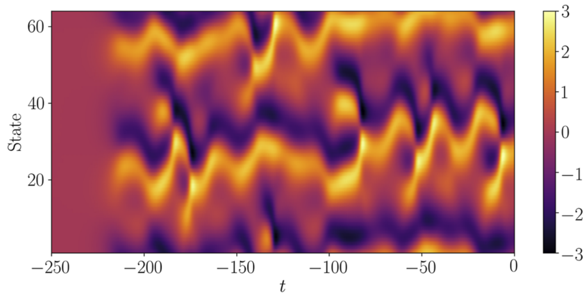

FIG. 1. The evolution of the KS system illustrating the spatiotemporal field data at the initial transient period.

connected devices equipped with sensors might be a game changer in weather monitoring and prediction. In their recent report, O'Grady *et al.*[6] highlighted that the Weather Company utilizes data from over 250 000 personal weather stations. Moreover, Chapman, Bell, and Bell[7] discussed how the crowdsourcing data-driven modeling paradigm could take meteorological science to a new level using smart Netatmo weather stations. As more attention shifts to smart and connected *internet of things* devices, security and privacy implications of such smart weather stations have also been discussed.[8] Additionally, big data will come with its own challenges characterized by 10 Vs.[9] The 10 Vs imply large volume, velocity, variety, veracity, value, validity, variability, venue, vocabulary, and vagueness. Volume refers to the size of data, velocity refers to the data generation rate, variety refers to the data type, veracity refers to the data quality and accuracy, value refers to the data usefulness, validity refers to the data quality and governance, variability refers to the dynamic, evolving behavior in the data source, venue refers to the heterogeneous data from multiple sources, and vocabulary refers to the semantics describing data structure. Finally, vagueness refers to the confusion over the meaning of data and tools used. In the weather forecast and many other processes, we foresee that all these problems will have to be addressed.

To this end, in this article, we focus on the statistical learning part and introduce a distributed training approach to generate autoencoder models that are relevant to the nonlinear dimensionality reduction of spatiotemporally distributed datasets. We aim at exploring the feasibility of such a decentralized learning framework to model complex spatiotemporal systems in which local data samples are held in edge devices. The case handled here is relatively simple but that was completely intentional as it eases the communication and dissemination of the work to a larger audience. Specifically, we put forth a federated ML framework considering the Kuramoto–Sivashinsky (KS) system,[10,11] which is known for its irregular or chaotic behavior.

This system has been derived to describe diffusion-induced chaotic behavior in reaction systems,[12] hydrodynamic instabilities in laminar flames,[13] phase dynamics of nonlinear Alfvén waves,[14] as well as nonlinear saturation of fluctuation potentials in plasma physics.[15] Due to its systematic route to chaos, the KS system has attracted much attention recently to test the feasibility of emerging ML approaches specifically designed to capture complex spatiotemporal dynamics (see, for example, González-García *et al.*,[16] Pathak *et al.*,[17] Vlachas *et al.*,[18] Linot and Graham,[19] and Vlachas *et al.*[20]). The KS equation with L-periodic boundary conditions can be written as

$$\frac{\partial u}{\partial t} = -\frac{\partial^4 u}{\partial x^4} - \frac{\partial^2 u}{\partial x^2} - u\frac{\partial u}{\partial x} \tag{1}$$

on a spatial domain $x \in [0, L]$, where the dynamics undergo a hierarchy of bifurcations as the spatial domain size L is increased, building up the chaotic behavior. Here, we perform the underlying numerical experiments with $L = 22$ to generate our spatiotemporal dataset. Equation (1) is solved using the fourth-order method for stiff partial

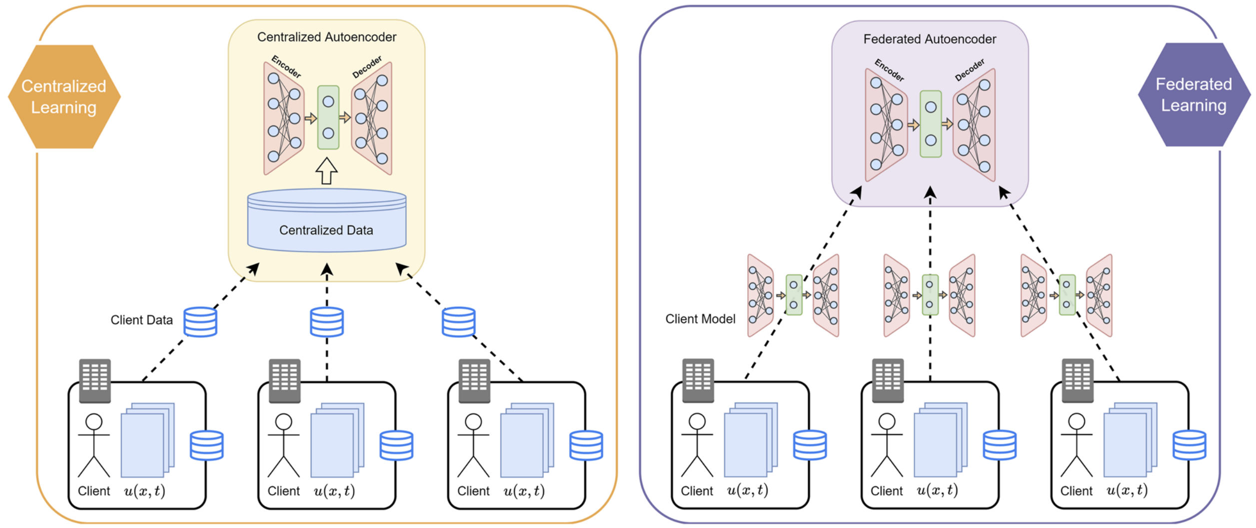

FIG. 2. Overview and schematic illustrations of the centralized and federated ML approaches.

differential equations[21] with the spatial grid size of $N = 64$. The random initial condition is assigned at time $t = -250$ and the solution is evolved with a time step of 2.5×10^{-3} up to $t = 0$. The trajectory of the KS system in the initial transient period is shown in Fig. 1. Using the solution at time $t = 0$ as the initial condition, the KS system is evolved until $t = 2500$. The data are sampled at a time step of 0.25, and these 10 000 samples are used for training and validation. For the testing purpose, the data from $t = 2500$ to $t = 3750$ are utilized.

FEDERATED MACHINE LEARNING

In this work, the federated ML is demonstrated for an autoencoder, which is a powerful approach for obtaining the latent space on a nonlinear manifold. The autoencoder is composed of the encoder and a decoder, where the encoder maps an input to a low-dimensional latent space and the decoder performs the inverse mapping from latent space variables to the original dimension at the output. If we denote the encoder function as $\eta(w)$ and a decoder function is defined as $\xi(w)$, we can represent the manifold learning as follows:

$$\eta, \xi = \underset{\eta,\xi}{\arg\max} \; \|\boldsymbol{u} - (\eta \circ \xi)\boldsymbol{u}\|, \tag{2}$$

$$\eta : \boldsymbol{u} \in \mathbb{R}^N \to \boldsymbol{z} \in \mathbb{R}^R, \tag{3}$$

$$\xi : \boldsymbol{z} \in \mathbb{R}^R \to \boldsymbol{u} \in \mathbb{R}^N, \tag{4}$$

where $\boldsymbol{z}$ represent the low dimensional latent space and R is the dimensionality of the latent space.

We closely follow the seminal work in federated learning,[22] which introduces a federated averaging algorithm where clients collaboratively train a shared model. Figure 2 contrasts the federated learning approach with the centralized method. In the centralized method, the local dataset is transferred from clients to a central server and the model is trained using centrally stored data. In case of the federated learning, the local dataset is never transferred from clients to a server. Instead, each client computes an update to the global model maintained by the server based on the local dataset, and only this update to the model is communicated. The federated averaging algorithm assumes that there is a fixed set of K clients with a fixed local dataset and a synchronous update scheme is applied in rounds of communications. At the beginning of each communication round, the central server sends the global state of the model (i.e., the current model parameters) to each of the clients. Each client computes the update to the global model based on the global state and local dataset and this update is sent to a server. The server then updates the global state of the model based on the local updates received from all clients, and this process continues. The objective

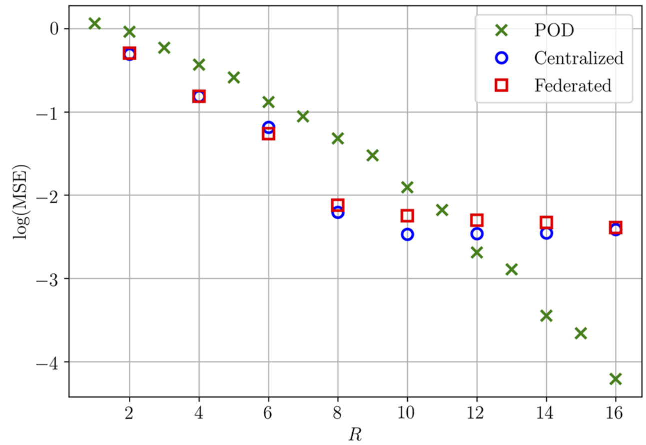

FIG. 4. Reconstruction mean squared error (MSE) on the test data.

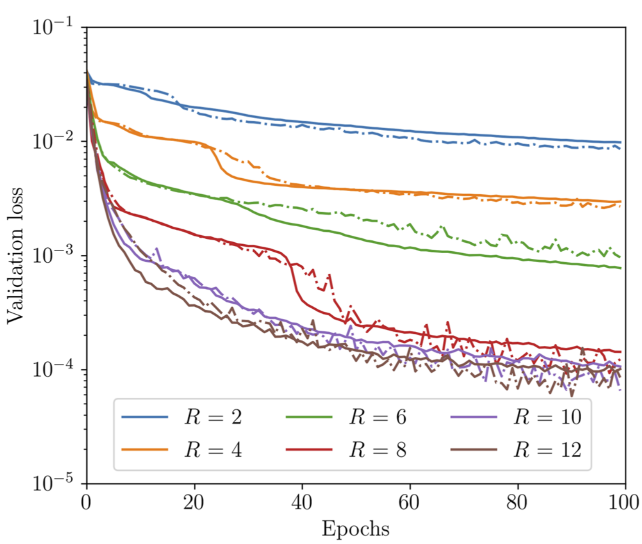

FIG. 3. Validation loss during training. Here, dashed line corresponds to centralized learning and solid lines are for federated learning.

ALGORITHM 1. Federated averaging algorithm. B is the local minibatch size, E is the number of local epochs, and α is the learning rate.

Server execution:
  initialize w_0
  for $t = 1, 2, \ldots$ **do**
    for each client k **do**
      $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
    end for
    $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$
  end for

ClientUpdate(k, w):
  $\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
  for each local epoch i from 1 to E **do**
    for batch $b \in \mathcal{B}$ **do**
      $w \leftarrow w - \alpha \nabla l(w; b)$
    end for
  end for
  return w to a server

function for a federated averaging algorithm can be written as follows:

$$f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \quad \text{where} \quad F_k(w) = \frac{1}{n_k} \sum_{i \in \mathcal{P}_k} f_i(w). \tag{5}$$

$\mathcal{P}_k$ is the data on the kth client, n_k is the cardinality of $\mathcal{P}_k$, and $f_i(w) = l(x_i, y_i;\ w)$ is the loss of the prediction on example (x_i, y_i). The above aggregation protocol can be applied to any ML algorithm. In this work, we use the autoencoder for nonlinear dimensionality reduction,[23] and the complete pseudo-code for deep learning models in a federated setting is provided in Algorithm 1. We highlight that the approach we utilize in our study simply weights edge devices proportionally by the data they own. More advanced approaches can be considered to mitigate such limitations,[24–28] but that is beyond the scope of this article.

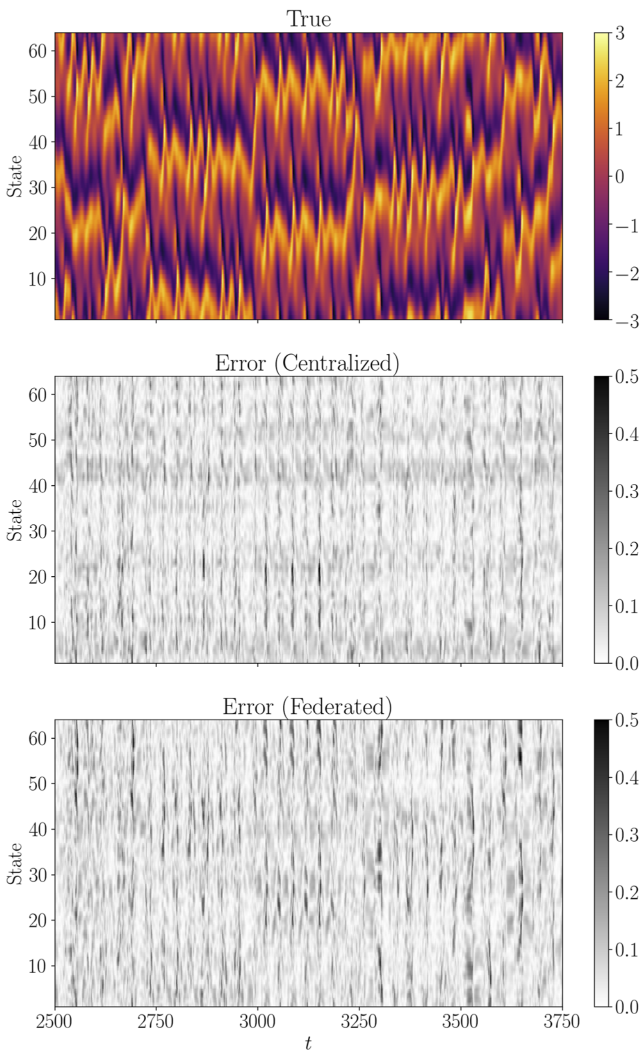

FIG. 5. Reconstruction performance of the centralized and federated learning approaches with $R = 8$.

Following the work of Vlachas *et al.*,[20] we first validate the centralized approach by varying R. For the federated learning, we use $K = 10$ clients, and each client model is trained for $E = 1$ local epoch with a batch size $B = 32$. For a fair comparison, the batch size of 320 is utilized for training the centralized autoencoder. The validation loss for the centralized and federated autoencoder with a different dimensionality of the latent space is depicted in Fig. 3, and we see that both the losses converge to very similar values. This shows that there is no significant loss in accuracy due to federated learning compared to centralized learning. As shown in Fig. 4, the reconstruction error for both centralized and federated autoencoders saturates around $R = 8$ modes. Figure 4 also demonstrates that a linear approach based on the proper orthogonal decomposition (POD) (see, e.g., the work of Ahmed *et al.*,[1,29] Pawar *et al.*,[30,31] and San and Iliescu[32,33]) requires significantly more modes to represent underlying flow dynamics with the same accuracy. Our observations, which are consistent with previous studies,[19,20,34,35] suggest that the latent space dynamics lies effectively on a manifold with $R = 8$ dimensions. Although our analysis includes a global POD approach for comparison purposes, we may consider to apply a localized POD approach[36–38] for improved modal representation. Instead of a detailed POD analysis here, our work rather aims primarily at demonstrating the potential of federated learning in fluid mechanics as opposed to centralized learning.

The trajectory of the KS system for the testing period is shown in Fig. 5 along with the error between the true data and reconstructed data from centralized and federated autoencoders. The error is computed as the absolute difference between the true and predicted state of the KS system. Both the centralized and federated autoencoders have a similar level of error.

CONCLUSION

This article explores the potential of federated ML for modeling complex spatiotemporal dynamical systems. In particular, we considered the problem of nonlinear dimensionality reduction of chaotic systems as a demonstration case. Federated learning allows for collaborative training of a model while keeping the training data decentralized. Our numerical experiments with the application of an autoencoder to the Kuramoto–Sivashinsky system show that a federated model can achieve the same level of accuracy as the model trained using the central data collected from all clients. This work opens up the possibility of updating a model in a centralized setting without exposing the local data collected from different sources.

We argue that federated learning can solve some of the *big data* challenges in complex dynamical systems provided that the different stakeholders, clients, and vendors use the same *vocabulary* as follows:

- *Big volume and velocity:* Since inference, analysis, and modeling happened only on the edge devices, a small amount of data needs to be communicated. This decentralizing process will significantly reduce the communication bandwidth and storage burden.
- *Big variety, venue, value, and vagueness:* Currently, a lack of trained personnel (to deal with a large variety of data in a centralized location) hinders the adoption of scalable digital

solutions. However, the problem is automatically remedied due to domain experts' presence at the data generation venue to extract value, thereby minimizing vagueness.
- *Big variability, veracity, and validity:* The variability in the data generation and sharing processes resulting from rapid changes in sensor technologies and corresponding regulatory environment will not be a challenge as it will be dealt with locally with federated learning.
- *Solving data privacy and security issues:* Since the data never leave the local servers, they will enhance security and encourage clients and vendors to collaborate.

Although, in this article, we primarily focus on federated learning in the context of spatiotemporal reconstruction of such chaotic systems, our approach can be generalized to large-scale computational settings beyond transport phenomena for which the research outcomes might improve broader modeling and simulation software capabilities to design cohesive, effective, and secure predictive tools for cross-domain simulations in the various levels of information density. In our future studies, we plan to leverage the decentralized learning approaches in the context of precision meteorology and develop new physics-guided federated learning approaches to forge new surrogate models compatible among heterogeneous computing environments.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, under Award No. DE-SC0019290. O.S. gratefully acknowledges the support from their Early Career Research Program.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Omer San: Conceptualization (lead); Funding acquisition (lead); Investigation (equal); Writing – original draft (equal). **Suraj Pawar**: Investigation (equal); Writing – original draft (equal). **Adil Rasheed**: Investigation (equal); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1] S. E. Ahmed, S. Pawar, O. San, A. Rasheed, T. Iliescu, and B. R. Noack, "On closures for reduced order models—A spectrum of first-principle to machine-learned avenues," Phys. Fluids **33**, 091301 (2021).

[2] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," IEEE Access **8**, 21980–22012 (2020).

[3] O. San, A. Rasheed, and T. Kvamsdal, "Hybrid analysis and modeling, eclecticism, and multifidelity computing toward digital twin revolution," GAMM-Mitt. **44**, e202100007 (2021).

[4] Report from the NASA machine learning workshop, April 17–19, 2018, Boulder, CO, https://esto.nasa.gov/wp-content/uploads/2020/03/2018MachineLearningWorkshop_Report.pdf.

[5] D. J. Mildrexler, M. Zhao, and S. W. Running, "A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests," J. Geophys. Res.: Biogeosci. **116**, G03025, https://doi.org/10.1029/2010jg001486 (2011).

[6] M. O'Grady, D. Langton, F. Salinari, P. Daly, and G. O'Hare, "Service design for climate-smart agriculture," Inf. Process. Agric. **8**, 328–340 (2021).

[7] L. Chapman, C. Bell, and S. Bell, "Can the crowdsourcing data paradigm take atmospheric science to a new level? A case study of the urban heat island of London quantified using Netatmo weather stations," Int. J. Climatol. **37**, 3597–3605 (2017).

[8] V. Sivaraman, H. H. Gharakheili, C. Fernandes, N. Clark, and T. Karliychuk, "Smart IoT devices in the home: Security and privacy implications," IEEE Technol. Soc. Mag. **37**, 71–79 (2018).

[9] F. N. Fote, S. Mahmoudi, A. Roukh, and S. A. Mahmoudi, "Big data storage and analysis for smart farming," in *2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech)* (IEEE, 2020), pp. 1–8.

[10] D. Armbruster, J. Guckenheimer, and P. Holmes, "Kuramoto–Sivashinsky dynamics on the center–unstable manifold," SIAM J. Appl. Math. **49**, 676–691 (1989).

[11] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, Cambridge, 2012).

[12] Y. Kuramoto, "Diffusion-induced chaos in reaction systems," Prog. Theor. Phys. Suppl. **64**, 346–367 (1978).

[13] G. I. Sivashinsky, "Nonlinear analysis of hydrodynamic instability in laminar flames—I. Derivation of basic equations," Acta Astronaut. **4**, 1177–1206 (1977).

[14] E. L. Rempel, A. C.-L. Chian, A. J. Preto, and S. Stephany, "Intermittent chaos driven by nonlinear Alfvén waves," Nonlinear Process. Geophys. **11**, 691–700 (2004).

[15] R. E. LaQuey, S. M. Mahajan, P. H. Rutherford, and W. M. Tang, "Nonlinear saturation of the trapped-ion mode," Phys. Rev. Lett. **34**, 391 (1975).

[16] R. González-García, R. Rico-Martínez, and I. G. Kevrekidis, "Identification of distributed parameter systems: A neural net based approach," Comput. Chem. Eng. **22**, S965–S968 (1998).

[17] J. Pathak, B. Hunt, M. Girvan, Z. Lu, and E. Ott, "Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach," Phys. Rev. Lett. **120**, 024102 (2018).

[18] P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Koumoutsakos, "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks," Proc. R. Soc. A **474**, 20170844 (2018).

[19] A. J. Linot and M. D. Graham, "Deep learning to discover and predict dynamics on an inertial manifold," Phys. Rev. E **101**, 062209 (2020).

[20] P. R. Vlachas, G. Arampatzis, C. Uhler, and P. Koumoutsakos, "Multiscale simulations of complex systems by learning their effective dynamics," Nat. Mach. Intell. **4**, 359–366 (2022).

[21] A.-K. Kassam and L. N. Trefethen, "Fourth-order time-stepping for stiff PDEs," SIAM J. Sci. Comput. **26**, 1214–1233 (2005).

[22] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics* (PMLR, 2017), pp. 1273–1282.

[23] S. E. Ahmed, O. San, A. Rasheed, and T. Iliescu, "Nonlinear proper orthogonal decomposition for convection-dominated flows," Phys. Fluids **33**, 121702 (2021).

[24] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," Proc. Mach. Learn. Syst. **2**, 429–450 (2020).

[25] T. Li, M. Sanjabi, A. Beirami, and V. Smith, "Fair resource allocation in federated learning," arXiv:1905.10497 (2019).

[26]A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning: A meta-learning approach," arXiv:2002.07948 (2020).
[27]Y. Deng, M. M. Kamani, and M. Mahdavi, "Adaptive personalized federated learning," arXiv:2003.13461 (2020).
[28]A. Z. Tan, H. Yu, L. Cui, and Q. Yang, "Towards personalized federated learning," IEEE Trans. Neural Networks Learn. Syst. (published online, 2022).
[29]S. E. Ahmed, S. M. Rahman, O. San, A. Rasheed, and I. M. Navon, "Memory embedded non-intrusive reduced order modeling of non-ergodic flows," Phys. Fluids **31**, 126602 (2019).
[30]S. Pawar, O. San, A. Nair, A. Rasheed, and T. Kvamsdal, "Model fusion with physics-guided machine learning: Projection-based reduced-order modeling," Phys. Fluids **33**, 067123 (2021).
[31]S. Pawar, S. E. Ahmed, O. San, and A. Rasheed, "Data-driven recovery of hidden physics in reduced order modeling of fluid flows," Phys. Fluids **32**, 036602 (2020).
[32]O. San and T. Iliescu, "Proper orthogonal decomposition closure models for fluid flows: Burgers equation," Int. J. Numer. Anal. Model., Ser. B **5**, 217–237 (2014).
[33]O. San and T. Iliescu, "A stabilized proper orthogonal decomposition reduced-order model for large scale quasigeostrophic ocean circulation," Adv. Comput. Math. **41**, 1289–1319 (2015).
[34]P. Cvitanović, R. L. Davidchack, and E. Siminos, "On the state space geometry of the Kuramoto–Sivashinsky flow in a periodic domain," SIAM J. Appl. Dyn. Syst. **9**, 1–33 (2010).
[35]J. C. Robinson, "Inertial manifolds for the Kuramoto-Sivashinsky equation," Phys. Lett. A **184**, 190–193 (1994).
[36]G. Tadmor, D. Bissex, B. Noack, M. Morzynski, T. Colonius, and K. Taira, "Fast approximated POD for a flat plate benchmark with a time varying angle of attack," in *4th Flow Control Conference* (AIAA, 2008), p. 4191.
[37]O. San and J. Borggaard, "Principal interval decomposition framework for POD reduced-order modeling of convective Boussinesq flows," Int. J. Numer. Methods Fluids **78**, 37–62 (2015).
[38]M. Ahmed and O. San, "Stabilized principal interval decomposition method for model reduction of nonlinear convective systems with moving shocks," Comput. Appl. Math. **37**, 6870–6902 (2018).