Reconstructing dissipative dynamical systems from spatially and temporally sparse sensors

Alex Guo
Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545

Galen T. Craven
Theoretical Division
Los Alamos National Laboratory
Los Alamos, NM 87545

Javier E. Santos
Earth and Environmental Sciences Division
Los Alamos National Laboratory
Los Alamos, NM 87545

Charles D. Young *
Center for Nonlinear Studies
Los Alamos National Laboratory
Los Alamos, NM 87545

Abstract

The reconstruction of fields from sensor data is a common task in scientific applications. In many cases, the observation is under resolved, and memory must be incorporated to uniquely recover the underlying system. For dissipative systems, dynamical systems theory provides guidance in formulating these data-driven reconstructions, which has been the focus of many optimization and machine learning approaches. Most models are restricted to data sampled at fixed positions and regular time intervals. We introduce a model which overcomes these limitations using attention mechanisms with spatial and temporal encodings. Our model is based on the Senseiver, which reconstructs fields from instantaneous sparse sensor measurements. Informed by time delay embedding theorems, we formulate an attention-based model that learns from sensor data at varying spatial position and sampling rates. We evaluate the model on systems exhibiting a limit cycle and spatiotemporal chaos.

1 Introduction

Inspired by success in reconstructing degraded images, neural networks (NNs) have been applied to system reconstruction in many scientific applications [22, 2, 15, 6, 18]. These problems are related to phase space reconstruction via Takens embedding theorem [20], which proves that for dynamics on an attractor, a time delay embedding of a scalar observable has a diffeomorphism to the underlying system. Data-driven applications of these approaches emerged shortly after using local principal components analysis [5] and multi-layer perceptrons (MLPs) [16]. Progress in implementation of machine learning algorithms has led to significant advances in the speed and accuracy of reconstruction and forecasting models via reservoir computers (RCs) [13], convolutional NNs [6], Gaussian process regression [10], symbolic regression [1], long-short term memory [28, 24], and neural ODEs (NODEs) [14, 23, 26].

Despite this success, limited availability of data prevents direct application of most models. Data is often noisy, and sensors often drift in space or collect data intermittently. These pose challenges for models trained on clean data available at fixed positions and uniform intervals. The noise issue can be addressed by combining the forecast or reconstruction with model-free data assimilation methods [8]. Several approaches have addressed variable sensor positions and time sampling using NODEs and

*Correspondence should be addressed to cyoung@lanl.gov

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

implicit neural representations [25, 9]. However, these approaches tend to rely on hyperparameter tuning and lack an underlying theoretical motivation. Time delay embeddings have been extended to the case of irregularly sampled data by 'bundle embeddings' [19, 10], which essentially concatenate the sampling rate into the state representation.

Here, we develop a model for reconstructing the field values of a dynamical system with variable measurement intervals and sensor positions. The approach relies on the bundle embedding theorem [19], which implies the data should reside on a long-time 'inertial manifold' or 'strange attractor' of dimension $d_{\mathcal{M}}$ and embedding dimension $m \geq 2d_{\mathcal{M}} + 1$ to ensure the existence of the reconstruction map. Positional and temporal encodings are implemented in an attention-based autoencoder as an extension of the Senseiver model [17]. Compared to the previous model, the time delay embedding improves reconstructions when the number of sensors is lower than the underlying manifold dimension. For low Reynolds number flow past a cylinder—a limit cycle—we demonstrate accurate reconstruction from a single sensor for: 1) a range of time sampling rates and 2) variable sensor positions (both cases done separately). We also compare the model to previous work using MLPs [26] and RCs [13] for reconstructing the Kuramoto-Sivashinsky equation (KSE) in a chaotic parameter regime.

2 Methodology

We build upon the attention-based Senseiver model [17] by incorporating time delay embeddings. Let s_i be the sensor measurement recorded at position x_i. For N_s sensor values, we form the time delay embedding $\boldsymbol{s}_d(t) \in \mathbb{R}^{N_s \times m}$, where m is the number of time delays. The matrix component $\boldsymbol{s}_d(i, j, t) = s_i(t - \tau_{ij})$ contains the jth delay of the ith sensor value, where t_{ij} and τ_{ij} are the corresponding sensor measurement time and delay spacing, t is the leading sensor measurement time, and $\tau_{ij} = t - t_{ij}$. The sensor positions $\boldsymbol{x} = \{\boldsymbol{x}_i\}_{i=1}^{N_s}$ are encoded with sines and cosines [17] to $\boldsymbol{a}_x$, where each physical coordinate value (scalar) is mapped to a vector of length 32. Each component in the delay spacing matrix $\boldsymbol{\tau}$ is encoded with a time2vec (t2v) encoding [11] to get $\boldsymbol{a}_t$ (see Appendix A.1). This space-time encoding is fed as input along with the sensor value embedding $\boldsymbol{s}_d(t)$ to an attention-based encoder E that maps to a latent space $\boldsymbol{Z}$; then, an attention-based decoder D maps $\boldsymbol{Z}$ and a query position $\boldsymbol{x}_q$ to get the field value $\hat{s}(t)$. In mathematical terms, the model is as follows:

$$\boldsymbol{a}_x = P_E(\boldsymbol{x}), \tag{1}$$

$$\boldsymbol{Z}(t) = E(\boldsymbol{s}_d(t), \boldsymbol{a}_x, \boldsymbol{a}_t) = E(\boldsymbol{s}_d(t), P_E(x_s), \text{t2v}(\boldsymbol{\tau})), \tag{2}$$

$$\hat{s}(\boldsymbol{x}_q, t) = D(\boldsymbol{Z}(t), \boldsymbol{a}_q) = D(\boldsymbol{Z}(t), P_E(\boldsymbol{x}_q)). \tag{3}$$

Figure 1 shows a schematic of the model. We query the spatial domain point-by-point to get the fully reconstructed field $\hat{\boldsymbol{s}}$ at the current time t. The loss is $\langle \|s - \hat{s}\|^2 \rangle_t$, where the norm is performed over spatial points and the averaging is done over all training snapshots. In this study we use all spatial positions from the high resolution for training. However, the Senseiver output is pointwise, meaning the model can be trained with any level of downsampling in the ground truth data or choice of random seeding from a full field [7]. Additional details of the architecture is in Appendix A.2.

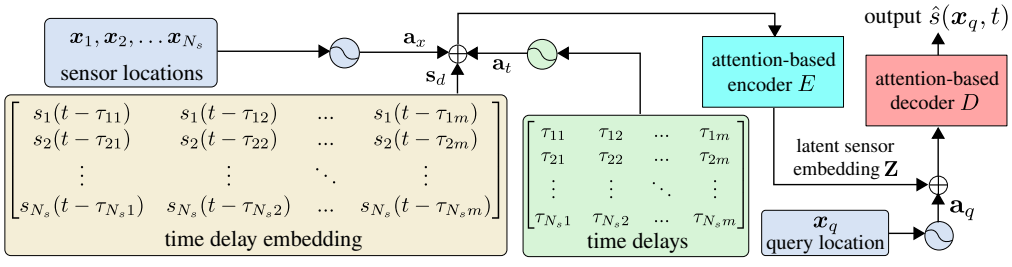

Figure 1: Architecture details. A time delay embedding of sensor values is concatenated with the positional and temporal encodings and encoded to a compressed latent space. The latent values are passed to the decoder with a query position to reconstruct the field at the current time.

3 Results

3.1 Non-uniform sampling in time

We first test the Senseiver on 2D flow past a cylinder at Re $= 100$, which settles onto a limit cycle with manifold dimension $d_{\mathcal{M}} = 1$. We reconstruct the steady state vorticity field ω from a single sensor with a variable sampling rate. The dataset consists of 5000 snapshots that cover ≈ 4 vortex shedding periods, of which 50 are used for training and the rest used for testing. During training and testing, the snapshots are variably spaced apart by 50 to 500 frames. As seen in Figure 2(d), the original Senseiver with no delay embeddings ($m = 1$) fails to capture the limit cycle, producing a mean L2 reconstruction error of $\epsilon = \|\omega - \hat{\omega}\|_2 / \|\omega\|_2 = 0.58$. This is expected, as we cannot prove the reconstruction map exists for $m < 2d_{\mathcal{M}} + 1$. Following Takens theorem, we select the embedding dimension $m = 3$. Concatenating the variable delay spacings τ_{0j} for $j = 1, 2$ to the $m = 3$ embedding of the sensor values yields a lower error of $\epsilon = 0.12$ (see Figure 2(e)). Upon using a t2v encoding of τ_{0j}, the error further decreases to $\epsilon = 0.05$ (see Figure 2(f)). As seen in Figure 2(c), the test error remains low for each pair of delay spacings τ_{0j} when using the method of t2v (while not shown, the heatmap plot of ϵ is similar for the other $m = 3$ method involving concatenation). We interpret the t2v encoding improvement as an increase of expressiveness of the model in predicting with a range of delay spacings rather than a single 'optimal' choice of delay coordinates.

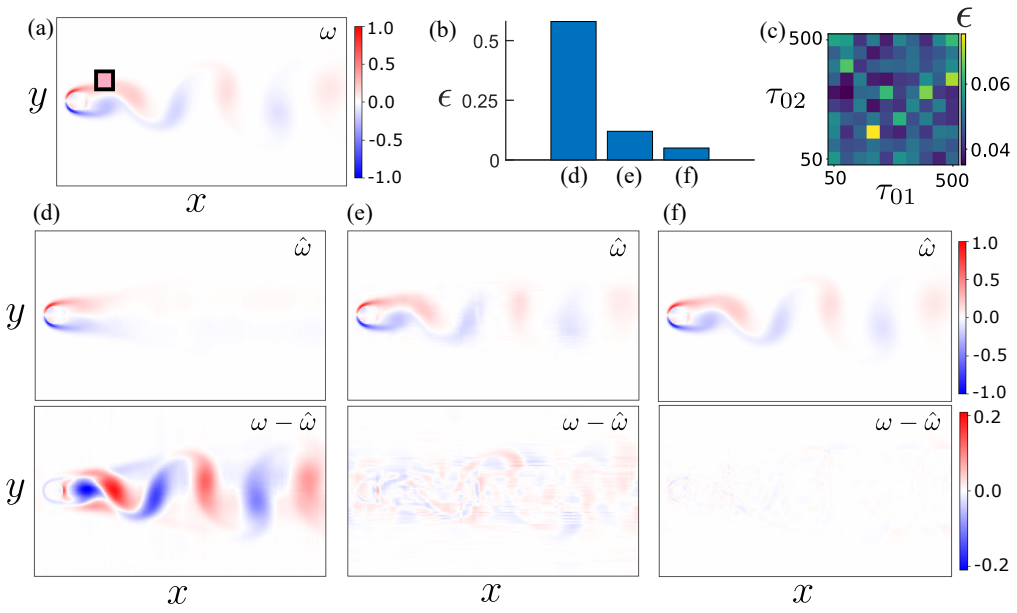

Figure 2: Field reconstruction of flow past a cylinder from sensor measurements that are non-uniform in time. (a) True vorticity field ω, with location of sensor marked with a square. (b) Reconstruction error for methods shown in panels (d)-(f). (c) Reconstruction error across each pair of delay spacings τ_{0j} for method shown in panel (f). Reconstruction fields $\hat{\omega}$ for models trained with (d) no time delay embedding, (e) embedding dimension $m = 3$ with temporal encoding of τ_{0j} via concatenation, and (f) $m = 3$ with temporal encoding of τ_{0j} via t2v are shown in the top panels of each subfigure. Difference in true and reconstructed fields are shown in the respective bottom panels of (d)-(f). In each field image, bounds on the y-axis are $[-9, 9]$ and the x-axis are $[0, 32]$ in dimensionless units.

3.2 Variable sensor position

We now test the model's robustness to sensor positions not seen during training. We work with uniformly sampled data here, so the positional encoding is still used but the temporal encoding is not. During training and testing, we use one sensor and its 3 time delays (spaced apart by 100 frames) as input to the model. 16 sensor positions in the vortex shedding region with high variance in the sensor values [6] are chosen for training. During testing, we individually evaluate the model for each and every point in the domain. As seen in Figure 3(a), the test error is generally lower for testing

locations at or near the 16 training locations (see Figures 3(b,c)). The high error at points exterior of the vortex-shedding region is due to the signals exhibiting small variance at all times. While the error is high for most out-of-training points interior of the vortex-shedding region, Figures 3(d,e) indicate that the model still reconstructs data on the true attractor but with a phase shift from the ground truth. The phase inaccuracy is likely a result of the chosen delay spacing not being optimal for all sensor positions.

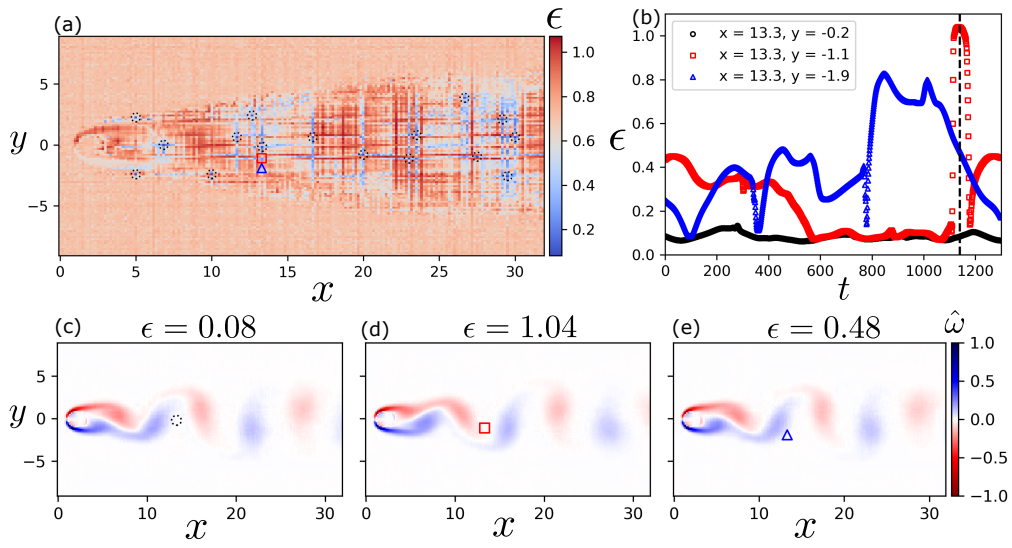

Figure 3: (a) Time and space averaged reconstruction test error for using a given sensor position as model input; dashed blue circles indicate training positions. (b) Error for each frame in a vortex shedding period for three sensor locations; vertical line corresponds to the snapshot time below for panels (c)-(e), which visualize the reconstruction at the selected locations. Panel (c) is a training position $x \approx 14, y \approx 0$, whereas (d) $y \approx -0.5$ and (e) $y \approx -1$ are test positions.

3.3 Chaotic systems

Lastly, we apply our model to the KSE, which is a 1D partial differential equation exhibiting diverse dynamical behavior. The domain size is $L = 22$ with periodic boundary conditions [3], for which there is sustained chaos on an inertial manifold of dimension $d_{\mathcal{M}} = 8$ [4, 12, 27]. We compare Senseiver to previous work using MLPs with time delays [26] and reservoir computing [13]. For all three methods, the training data size is kept constant at 60,000 with uniform time interval $\Delta t = 0.25$, the test data size is 10^5, and the reconstruction error ϵ is calculated for spatial points not corresponding to sensor positions. We examine the case of 1, 2, 4, and 8 sensors (which are fixed and evenly-spaced) and adjust m so that $N_s m = 16 = 2d_{\mathcal{M}}$ with $\tau = 8\Delta t$. We list ϵ in Table 1 for the three methods. For all but the 8-sensor case, our Senseiver model with time delays achieves a lower error than the other two methods. Moreover, our model does so with fewer parameters, which aids in computational efficiency during training and testing.

Our method also generalizes well to non-uniform time spacing (not shown), achieving an error of $\epsilon = 0.12$ for $N_s = 4$ and $m = 4$ delays with spacings ranging from $2\Delta t$ to $16\Delta t$. These results will reported in further detail along with a study of variable sensor positions in a forthcoming full publication.

Table 1: Reconstruction error for Kuramoto-Sivashinsky equation (KSE). For all methods except reservoir computing (RC), the number of delays m and sensors N_s are set such that $N_s m = 16$.

Method	$\epsilon\,(N_s = 1)$	$\epsilon\,(N_s = 2)$	$\epsilon\,(N_s = 4)$	$\epsilon\,(N_s = 8)$	parameters
Senseiver	**0.68**	**0.37**	**0.13**	0.05	$\mathbf{2 \times 10^4}$
MLP [Young et al. 2023]	0.70	0.44	0.15	**0.03**	3×10^5
RC [Lu et al. 2017]	0.91	0.68	0.31	0.10	2×10^5

4 Conclusion

We developed an attention-based model for reconstructing the fields of dissipative dynamical systems from sensor measurements that are spatially sparse and irregularly sampled in time. We expanded on the Senseiver [17] by allowing for time delay embeddings as the input. We first applied our model to flow past a cylinder, where we reconstructed the full vorticity field from a single sensor with irregular time sampling. The model achieved low reconstruction error by using a delay embedding dimension of 3 along with a time2vec encoding [11] of the delay spacings to account for the temporal intermittency of the sensor. Next, we tested the model's ability to reconstruct the vorticity field from variable sensor positions not seen during training. Facilitated by the spatial encoding of the sensor positions, the model achieved modest out-of-distribution generalization; however, there was some phase error in the reconstruction for test sensor positions that were not located near the training positions. Lastly, we applied the model to the spatiotemporally chaotic KSE for non-variable sensor positions and uniform time sampling. While our model achieved similar reconstruction performance to MLPs, the attention mechanism used in our model allows for the computational complexity (and thus number of parameters) to scale less strongly with the length of the space-time encoding, which can be quite long and expensive to process with MLPs. We have demonstrated the developed model is capable of reconstructing fields from spatially and temporally sparse sensors that also vary in location. Code to reproduce the results will be included in a forthcoming publication, and the code for the original Senseiver is available at https://github.com/OrchardLANL/Senseiver.

Acknowledgments and Disclosure of Funding

We acknowledge support from the Los Alamos National Laboratory (LANL) Directed Research and Development funds (LDRD). This research was performed in part at the Center for Nonlinear Studies (CNLS) at LANL. This research used resources provided by the Darwin testbed at LANL which is funded by the Computational Systems and Software Environments subprogram of LANL's Advanced Simulation and Computing program. LANL is operated by the Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (contract no. 89233218NCA000001).

References

[1] Joseph Bakarji, Kathleen Champion, J Nathan Kutz, and Steven L Brunton. Discovering governing equations from partial measurements with deep delay autoencoders. *Proceedings of the Royal Society A*, 479(2276):20230422, 2023.

[2] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019.

[3] Predrag Cvitanovic, Roberto Artuso, Ronnie Mainieri, Gregor Tanner, Gábor Vattay, Niall Whelan, and Andreas Wirzba. *Chaos: classical and quantum*. Niels Bohr Institute, Copenhagen, 2016.

[4] X Ding, H Chaté, P Cvitanović, E Siminos, and KA Takeuchi. Estimating the dimension of an inertial manifold from unstable periodic orbits. *Physical Review Letters*, 117(2):024101, 2016.

[5] J Doyne Farmer and John J Sidorowich. Predicting chaotic time series. *Physical review letters*, 59(8):845, 1987.

[6] Kai Fukami, Romit Maulik, Nesar Ramachandra, Koji Fukagata, and Kunihiko Taira. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nature Machine Intelligence 2021 3:11*, 3(11):945–951, 10 2021.

[7] Alejandro Güemes, Carlos Sanmiguel Vila, and Stefano Discetti. Super-resolution generative adversarial networks of randomly-seeded fields. *Nature Machine Intelligence*, 4(12):1165–1173, 2022.

[8] Franz Hamilton, Tyrus Berry, and Timothy Sauer. Ensemble kalman filtering without a model. *Physical Review X*, 6(1):011021, 2016.

[9] Valerii Iakovlev, Markus Heinonen, and Harri Lähdesmäki. Learning space-time continuous latent neural pdes from partially observed states. *Advances in Neural Information Processing Systems*, 36, 2024.

[10] Bethany Johnson and Stephan B Munch. An empirical dynamic modeling framework for missing or irregular samples. *Ecological Modelling*, 468:109948, 2022.

[11] Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2vec: Learning a vector representation of time. *arXiv*, 2019.

[12] Alec J Linot and Michael D Graham. Deep learning to discover and predict dynamics on an inertial manifold. *Physical Review E*, 101(6):062209, 2020.

[13] Zhixin Lu, Jaideep Pathak, Brian Hunt, Michelle Girvan, Roger Brockett, and Edward Ott. Reservoir observers: Model-free inference of unmeasured variables in chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(4):041102, 2017.

[14] Said Ouala, Duong Nguyen, Lucas Drumetz, Bertrand Chapron, Ananda Pascual, Fabrice Collard, Lucile Gaultier, and Ronan Fablet. Learning latent dynamics for partially observed chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(10):103121, 2020.

[15] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A. Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1):86–109, 2020.

[16] Ramiro Rico-Martinez, K Krischer, IG Kevrekidis, MC Kube, and JL Hudson. Discrete-vs. continuous-time nonlinear signal processing of cu electrodissolution data. *Chemical Engineering Communications*, 118(1):25–48, 1992.

[17] Javier E Santos, Zachary R Fox, Arvind Mohan, Daniel O'Malley, Hari Viswanathan, and Nicholas Lubbers. Development of the senseiver for efficient field reconstruction from sparse observations. *Nature Machine Intelligence*, 5(11):1317–1325, 2023.

[18] Daniel A Serino, Marc L Klasky, Balasubramanya T Nadiga, Xiaojian Xu, and Trevor Wilcox. Reconstructing richtmyer-meshkov instabilities from noisy radiographs using low dimensional features and attention-based neural networks. *arXiv preprint arXiv:2408.00985*, 2024.

[19] Jaroslav Stark, David S Broomhead, Michael Evan Davies, and J Huke. Takens embedding theorems for forced and stochastic systems. *Nonlinear Analysis: Theory, Methods & Applications*, 30(8):5303–5314, 1997.

[20] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *Advances in Neural Information Processing Systems*, 2017-Decem(Nips):5999–6009, 6 2017.

[22] Ge Wang, Jong Chu Ye, Klaus Mueller, and Jeffrey A. Fessler. Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging*, 37(6):1289–1296, 2018.

[23] Zhe Wang and Claude Guet. Reconstructing a dynamical system and forecasting time series by self-consistent deep learning. *arXiv preprint arXiv:2108.01862*, 2021.

[24] Jan P Williams, Olivia Zahn, and J Nathan Kutz. Sensing with shallow recurrent decoder networks. *Proceedings of the Royal Society A*, 480(2298):20240054, 2024.

[25] Yuan Yin, Matthieu Kirchmeyer, Jean-Yves Franceschi, Alain Rakotomamonjy, and Patrick Gallinari. Continuous pde dynamics forecasting with implicit neural representations. *arXiv preprint arXiv:2209.14855*, 2022.

[26] Charles D Young and Michael D Graham. Deep learning delay coordinate dynamics for chaotic attractors from partial observable data. *Physical Review E*, 107(3):034215, 2023.

[27] Kevin Zeng, Carlos E Pérez De Jesús, Andrew J Fox, and Michael D Graham. Autoencoders for discovering manifold dimension and coordinates in data from complex dynamical systems. *Machine Learning: Science and Technology*, 5(2):025053, 2024.

[28] Elise Özalp, Georgios Margazoglou, and Luca Magri. Reconstruction, forecasting, and stability of chaotic dynamics from partial data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(9), September 2023.

A Appendix / supplemental material

A.1 Time2vec encoding

The time2vec encoding [11] maps a scalar τ to a vector

$$\text{t2v}(\tau)[k] = \begin{cases} \omega_k \tau + b_k \,, & \text{if } k = 0 \\ F(\omega_k \tau + b_k) \,, & \text{for } 1 \leq k \leq K-1 \end{cases},$$

where $\text{t2v}(\tau)[k]$ is the kth component of $\text{t2v}(\tau) \in \mathbb{R}^K$, F is an activation function, and ω_k and b_k are learnable parameters. We choose F to be the sine function and K to be 32. The t2v encoding allows both cyclic and non-cyclic information to be encoded.

A.2 Architecture details

Here we provide additional details on the attention-based encoder and decoder blocks in Figure 1, which are based on the architecture in the original Senseiver model [17]. We use the scaled dot product attention mechanism [21]

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\mathrm{T}}}{\sqrt{d_m}}\right)\mathbf{V}, \tag{4}$$

where d_m is the model dimension of query matrix $\mathbf{Q}$, key matrix $\mathbf{K}$, and value matrix $\mathbf{V}$.

Multi-head attention applies a set of N_H linear transformations on $\mathbf{Q}, \mathbf{K}$ and $\mathbf{V}$ and computes the attention mechanism on each linear transformation. Then, the outputs from each head are concatenated and passed through a final linear transformation to form the output:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = [\mathbf{O}_1 \oplus \cdots \oplus \mathbf{O}_{N_H}]\,\mathbf{W}_M \tag{5}$$

$$\text{where } \mathbf{O}_h = \text{Attention}(\mathbf{Q}\mathbf{W}_h^Q, \mathbf{K}\mathbf{W}_h^K, \mathbf{V}\mathbf{W}_h^V). \tag{6}$$

For the time-delayed Senseiver model, the sequence length is $d_s = N_s m$, since each sensor has m delays. The model dimension is $d_m = N_I + 2N_f N_d + K$, where N_I is the number of channels of the sensor measurements (one for the cases considered in this paper), N_f is the number of pairs of sine-cosine frequencies used in the positional encoding for each of the N_d number of spatial directions, and K is the length of the time2vec encoding.

The attention-based encoder processes the input $\mathbf{E}^{(0)} \in \mathbb{R}^{d_s \times d_m}$ with a linear layer to give an intermediate output $\mathbf{E}^{(1)} \in \mathbb{R}^{d_s \times N_c}$, where N_c is the number of latent channels. Next, a multi-head cross attention is performed between a trainable query array $\mathbf{Q}_{\text{in}} \in \mathbb{R}^{N_{Q_{\text{in}}} \times N_c}$ and key and value matrices derived from $\mathbf{E}^{(1)}$ to yield $\mathbf{E}^{(2)}$; here, $N_{Q_{\text{in}}}$ is the latent sequence length that is held constant to allow the attention calculations to be independent of N_s and m. $\mathbf{E}^{(2)}$ is then processed with multi-head self-attention to give the latent space $\mathbf{Z} \in \mathbb{R}^{N_{Q_{\text{in}}} \times N_c}$. The attention-based decoder performs similar steps as the encoder (see [17]), but includes a final MLP layer to reduce the dimension and obtain the predicted measurement $\hat{s}$ at the queried location x_q and current time t.