

---

# Harnessing Loss Decomposition for Long-Horizon Wave Predictions via Deep Neural Networks

---

**Indu Kant Deo**

Department of Mechanical Engineering  
The University of British Columbia  
Vancouver, BC V6T 1Z4  
indukant@mail.ubc.ca

**Rajeev K. Jaiman**

Department of Mechanical Engineering  
The University of British Columbia  
Vancouver, BC V6T 1Z4  
rjaiman@mail.ubc.ca

## Abstract

Accurate prediction over long time horizons is crucial for modeling complex physical processes such as wave propagation. Although deep neural networks show promise for real-time forecasting, they often struggle with accumulating phase and amplitude errors as predictions extend over a long period. To address this issue, we propose a novel loss decomposition strategy that breaks down the loss into separate phase and amplitude components. This technique improves the long-term prediction accuracy of neural networks in wave propagation tasks by explicitly accounting for numerical errors, improving stability, and reducing error accumulation over extended forecasts.

## 1 Introduction

Accurate long-horizon predictions are essential for understanding and modeling complex physical phenomena, particularly in the context of wave propagation governed by partial differential equations (PDEs) [17]. Traditional numerical methods for solving PDEs, such as finite difference, finite element, and spectral methods, have been extensively studied and applied [2] to solve the wave equation. Although these methods are effective, they often require substantial computational resources, making them unsuitable for real-time predictions and control related tasks.

Machine learning techniques, particularly deep neural networks, have emerged as powerful tools for modeling complex physical phenomena, including wave propagation and underwater noise prediction [6, 5, 9, 7, 8]. Despite their success in numerous domains, neural networks often struggle with accurate long-term predictions, primarily due to accumulated phase and amplitude errors [16, 11, 13]. This issue is particularly pronounced in autoregressive models, where errors can compound over time, leading to significant deviations from the true dynamics [20, 4].

Recent studies have highlighted the importance of loss decomposition in addressing these challenges. Traditional mean squared error (MSE) loss functions do not distinguish between different types of prediction errors, such as phase and amplitude errors, which are critical to accurate long-term forecasting [15, 21, 3]. To mitigate this, we propose decomposing the MSE into dissipation and dispersion errors [12], which separately account for the amplitude and phase errors, respectively. By explicitly considering these error types, our approach aims to enhance the neural network's ability to maintain accurate predictions over extended horizons.

The concept of loss decomposition is not entirely new; it has been previously explored in numerical analysis and signal processing. For example, Kreiss and Olinger [14] introduced the idea of dissipation and dispersion errors in the context of finite difference methods. More recently, Guen et al. [15] applied similar ideas to machine learning, suggesting that decomposed loss functions can guide neural networks to learn more robust representations of time series forecasting. Our work builds on

these ideas by specifically targeting the phase and amplitude errors that hinder long-term predictions in neural networks for wave propagation tasks. By integrating the perspective of numerical errors into machine learning techniques, we introduce a novel theoretical framework aimed at improving long-term predictions.

## 2 Methods

### 2.1 Linear advection equation

In this study, we focus on data generated from the linear convection equation as a model problem for wave propagation. Although the approach has been applied to the inviscid Burgers equation to investigate nonlinear wave propagation, we limit our discussion to the linear convection equation for brevity. The solution  $U$  in the domain  $\Omega \equiv [0, 1]$  satisfies the following parametric partial differential equation:

$$\frac{\partial U}{\partial t} + \mu \frac{\partial U}{\partial X} = 0 \quad \text{in } \Omega, \quad (1)$$

with the initial condition:

$$U(x, 0) = U_0(x) \equiv f(x), \quad (2)$$

where  $\mu \in [0.775, 1.25]$  represents the wave speed. The initial profile is given by  $f(x) = \left(\frac{1}{\sqrt{2\pi\sigma}}\right) e^{-x^2/2\sigma}$  with  $\sigma = 5 \times 10^{-3}$ . A one-dimensional spatial discretization of 256 grid points and 200 time steps is applied. The training set consists of  $N_\mu = 20$  parameter instances, uniformly distributed over the range of  $\mu$ , while the test set includes  $N_{\text{test}} = 19$  parameter instances. The test parameters are selected such that  $\mu_{\text{test},i} = \frac{\mu_{\text{train},i} + \mu_{\text{train},i+1}}{2}$  for  $i = 1, \dots, N_{\text{test}}$ . A visual representation of this dataset is provided in Figure 1.

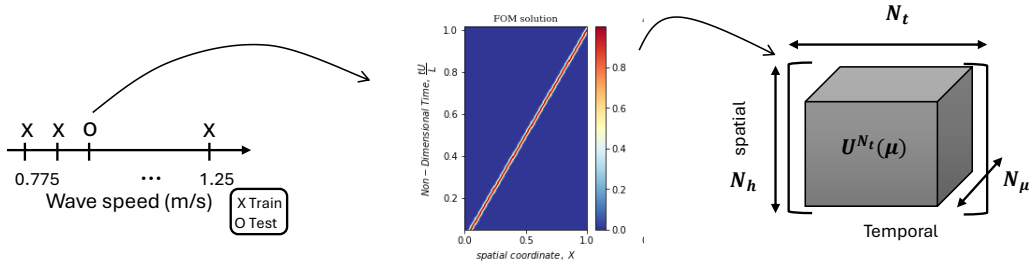


Figure 1: Schematic of dataset generation. Linear convection simulations were generated for parameter wave speed. Amplitude data were extracted and arranged into a training snapshot matrix.

### 2.2 Attention-based Convolution Recurrent Autoencoder Network

We employ a deep learning architecture to model wave propagation, specifically utilizing a 15-layer Attention-Based Convolutional Recurrent Autoencoder Network [6, 10]. The encoder includes a convolutional layer, followed by max pooling and fully connected layers. The output layer of the encoder contains  $r$  neurons, where  $r$  represents the dimensionality of the reduced manifold. The propagator is composed of attention-based sequence-to-sequence RNN-LSTM layers that process the reduced manifold data, advancing it along the time dimension. The decoder, designed as the reverse of the encoder, reconstructs the data from the latent space back to the original physical dimension. The model architecture and hyperparameters were adopted from the work of Deo et al. [6]. The model is trained from scratch using TensorFlow [1] on a single NVIDIA P2200 Pascal GPU, with training converging in 285 epochs within 20 minutes of wall-clock time. Figure 2 shows the visual representation of AB-CRAN architecture.

### 2.3 Loss decomposition

Mean squared error (MSE) is commonly employed as the default loss function for regression tasks. However, MSE fails to account for numerical dissipation and dispersion errors in the solution. To

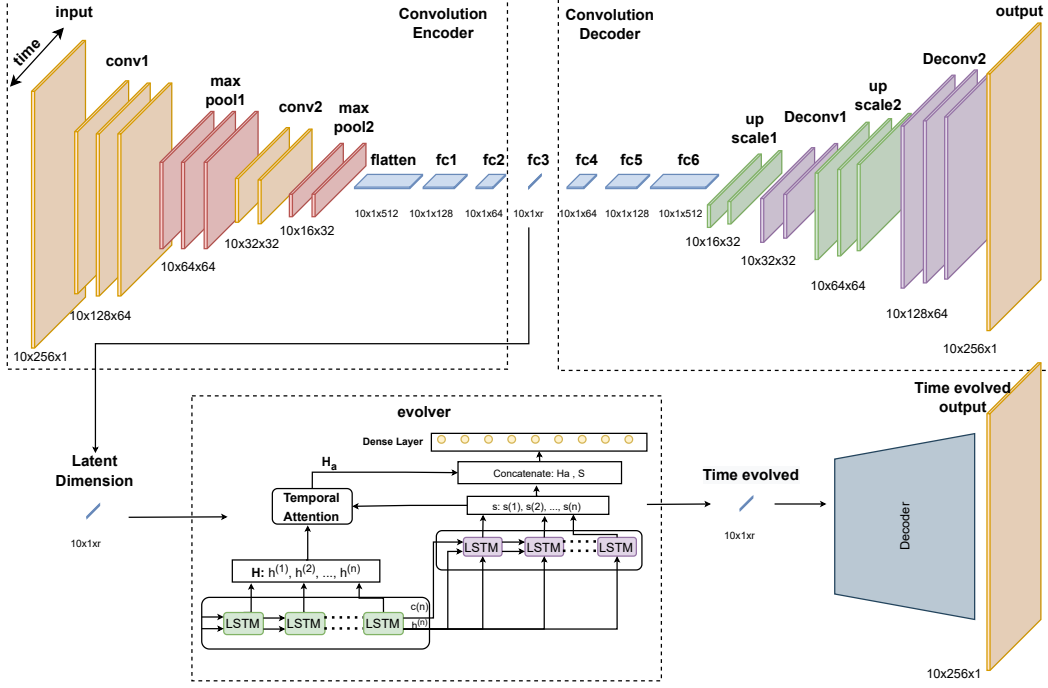


Figure 2: Illustration of attention-based convolutional recurrent autoencoder architecture.

address this limitation, MSE can be decomposed into two components: one representing dissipation error (amplitude) and the other capturing dispersion error (phase). The total mean square error can be rewritten as:

$$\tau = \sigma^2(u_a) + \sigma^2(u_d) - 2\rho\sigma(u_a)\sigma(u_d) + (\bar{u}_a - \bar{u}_d)^2, \quad (3)$$

where  $u_a$  is the ground-truth solution and  $u_d$  the predicted solution.  $\sigma^2(u_a)$ ,  $\sigma^2(u_d)$  are the variances of  $u_a$  and  $u_d$ , respectively.  $\text{cov}(u_a, u_d)$  is the covariance between  $u_a$  and  $u_d$  and  $\rho$  is the correlation coefficient. Eq. 3 can be rewritten as

$$\tau = [\sigma(u_a) - \sigma(u_d)]^2 + (\bar{u}_a - \bar{u}_d)^2 + 2(1 - \rho)\sigma(u_a)\sigma(u_d). \quad (4)$$

We define the first two terms of the RHS of Eq. 4 as the dissipation error and the third term as the dispersion error:

$$\begin{aligned} \tau_{\text{DISS}} &= [\sigma(u_a) - \sigma(u_d)]^2 + (\bar{u}_a - \bar{u}_d)^2, \\ \tau_{\text{DISP}} &= 2(1 - \rho)\sigma(u_a)\sigma(u_d). \end{aligned} \quad (5)$$

When two wave patterns differ solely in amplitude but share the same phase, their correlation coefficient,  $\rho$ , should be 1, with  $\tau_{\text{DISP}} = 0$  as indicated by equation (5). This outcome aligns with our expectations and is a reasonable result. We employ this decomposition to formulate the loss function as follows:

$$\begin{aligned} \mathcal{L}_{\text{function}} &= (1 - \alpha)\mathcal{L}_{\text{Decoder}} + \alpha\mathcal{L}_{\text{propagator}}, \\ \mathcal{L}_{\text{propagator}} &= (1 - \beta)\tau_{\text{DISP}} + \beta\tau_{\text{DISS}}, \end{aligned} \quad (6)$$

where  $\alpha$  and  $\beta$  are hyperparameters. To determine the values of  $\alpha$  and  $\beta$ , we utilize hyperparameter tuning with the Ray Tune ASHA algorithm [18]. Our experiments reveal that setting  $\alpha$  and  $\beta$  greater than 0.5 prioritizing the propagator, with a stronger emphasis on reducing dispersion error within the propagator works best. This work presents the first introduction of phase and amplitude components into the loss function for training deep neural networks for wave propagation.

## 2.4 Training

To improve the training of the attention-based convolutional recurrent autoencoder (AB-CRAN), we integrate a denoising mechanism into the decoder's loss function. During the initial forward

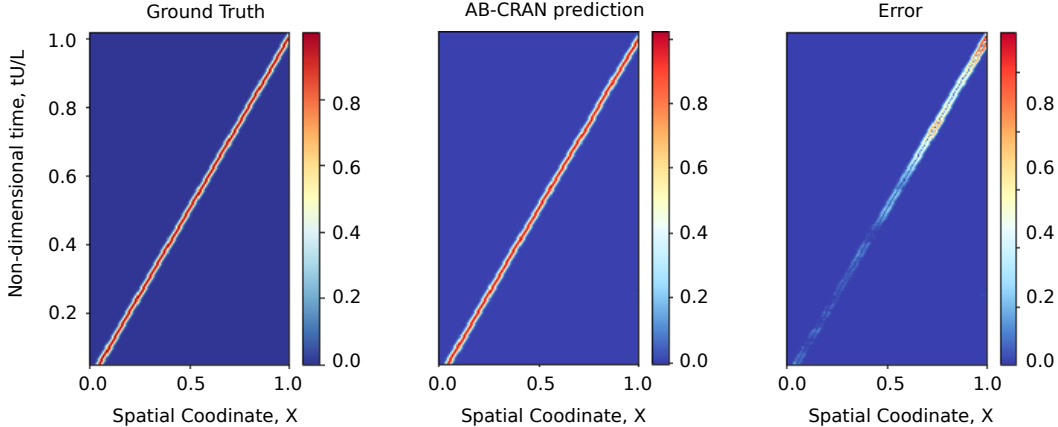


Figure 3: Linear convection problem: Exact solution (left), AB-CRAN solution with  $n = 2$  (center) and error  $e = |\hat{u} - u|$  (right) for the testing parameter  $\mu_{test} = 1.0125$  in the space-time domain.

pass, noise is added to the input data before passing it through the encoder-decoder pipeline. The decoder is then tasked with reconstructing the original noise-free data, optimizing the mean squared error between the noisy inputs and the clean data, resulting in a denoising decoder loss. In the second stage, the projected data from the encoder are processed through the propagator to generate a time-evolved output, and the corresponding loss is decomposed into dissipation and dispersion errors, as explained in the above section. The entire model is optimized using the AdamW optimizer [22] and a cosine annealing warm-restart scheduler [19]. We also implemented an early stopping strategy with a patience threshold of 50 epochs.

### 3 Results

We randomly selected a parameter from the test set for evaluation. The AB-CRAN architecture was provided with data from the linear convection equation, using  $\mu = 1.0125$  as input. Figure 3 illustrates the exact solution alongside the AB-CRAN approximation for this test case. The results demonstrate that the AB-CRAN model, trained with the proposed loss decomposition strategy, captures wave dynamics with high fidelity, accurately preserving key features of the solution.

Our loss decomposition strategy significantly improves phase accuracy compared to the conventional mean squared error (MSE) loss. As shown in Figure 4a, predictions based on MSE exhibit noticeable phase lag over time, highlighting the inability of traditional loss functions to maintain phase alignment. In contrast, our method preserves phase throughout the prediction horizon. Furthermore, Figure 4b demonstrates that the proposed approach extends the prediction horizon by effectively mitigating error accumulation. By separately addressing phase and amplitude errors, the model achieves greater stability and accuracy, particularly for long-term forecasts.

This decomposition-based approach provides a robust and generalizable framework for long-term forecasting of physical systems. By isolating and addressing phase and amplitude errors, it overcomes limitations inherent in traditional MSE-based methods. The strategy reduces error accumulation, ensuring phase alignment and amplitude stability over extended time horizons. These improvements make the method broadly applicable to various physical forecasting tasks, such as wave propagation, fluid dynamics, and climate modeling, establishing a new standard for accuracy and reliability in long-horizon predictions.

### 4 Conclusion

In this work, we presented a novel loss decomposition strategy that addresses the limitations of traditional mean squared error (MSE)-based loss functions for long-term predictions in physical systems. By separating the loss into phase and amplitude components, our approach independently tackles dispersion errors (phase misalignment) and dissipation errors (amplitude inaccuracies). This targeted decomposition significantly enhances both the accuracy and stability of predictions, extending

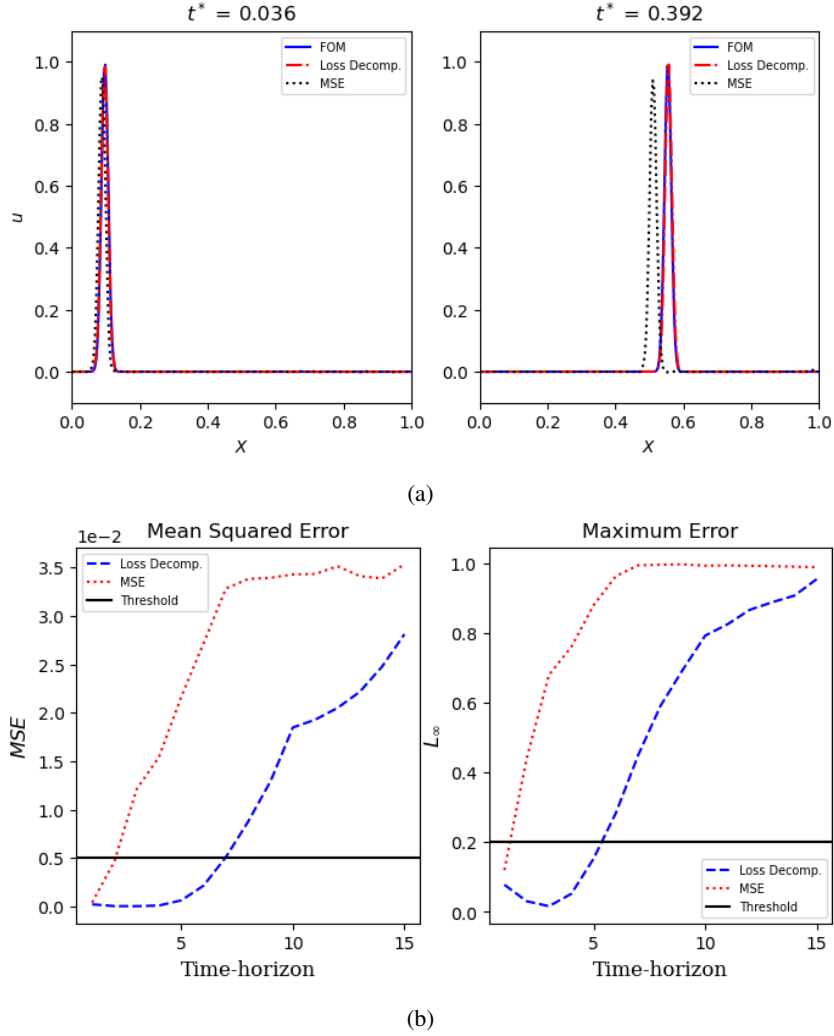


Figure 4: Accuracy assessment for wave prediction: (a) Comparison of MSE-based loss and Loss-decomposition-based training prediction for  $t^* = 0.036, 0.392$  (b) Error vs. time-horizon plot comparing MSE-based and loss-decomposition-based training predictions, with a time-horizon of 10 time-steps

the prediction horizon while maintaining physical consistency. Our results demonstrate that the proposed method outperforms standard MSE-based approaches, offering substantial improvements for tasks such as wave propagation modeling.

The flexibility of this framework makes it broadly applicable to various forecasting tasks in physical domains, including flow dynamics, ocean acoustics, and climate modeling. By addressing phase and amplitude errors independently, our method ensures reliable long-term predictions for complex systems. This study underscores the potential of tailored loss functions to advance the modeling and simulation of complex dynamical systems, providing a promising foundation for future research in data-driven physical modeling and scientific computing.

## References

- [1] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [2] John David Anderson and John Wendt. *Computational fluid dynamics*, volume 206. Springer, 1995.

- [3] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.
- [4] Tom Beucler, Michael Pritchard, Stephan Rasp, Jordan Ott, Pierre Baldi, and Pierre Gentine. Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, 126(9):098302, 2021.
- [5] Indu Kant Deo, Rui Gao, and Rajeev Jaiman. Combined space–time reduced-order model with three-dimensional deep convolution for extrapolating fluid dynamics. *Physics of Fluids*, 35(4), 2023.
- [6] Indu Kant Deo and Rajeev Jaiman. Predicting waves in fluids with deep neural network. *Physics of Fluids*, 34(6), 2022.
- [7] Indu Kant Deo, Akash Vankateshwaran, and Rajeev Jaiman. Predicting wave propagation for varying bathymetry using conditional convolutional autoencoder network. In *International Conference on Offshore Mechanics and Arctic Engineering*, volume 87844, page V006T08A038. American Society of Mechanical Engineers, 2024.
- [8] Indu Kant Deo, Akash Venkateshwaran, and Rajeev K Jaiman. Continual learning of range-dependent transmission loss for underwater acoustic using conditional convolutional neural net. *arXiv preprint arXiv:2404.08091*, 2024.
- [9] Rui Gao, Indu Kant Deo, and Rajeev K Jaiman. A finite element-inspired hypergraph neural network: Application to fluid dynamics simulations. *Journal of Computational Physics*, 504:112866, 2024.
- [10] Francisco J Gonzalez and Maciej Balajewicz. Deep convolutional recurrent autoencoders for learning low-dimensional feature dynamics of fluid systems. *arXiv preprint arXiv:1808.01346*, 2018.
- [11] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [12] Graham W. Griffiths. *Dissipation and Dispersion*, page 264–284. Cambridge University Press, 2016.
- [13] Anuj Karpatne, Gowtham Atluri, James H Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on knowledge and data engineering*, 29(10):2318–2331, 2017.
- [14] Heinz-Otto Kreiss and Joseph Oliger. Comparison of accurate methods for the integration of hyperbolic equations. *Tellus*, 24:199–215, 1972.
- [15] Vincent Le Guen and Nicolas Thome. Shape and time distortion loss for training deep time series forecasting models. *Advances in neural information processing systems*, 32, 2019.
- [16] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.
- [17] R. J. LeVeque et al. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.
- [18] L Li, K Jamieson, A Rostamizadeh, E Gonina, M Hardt, B Recht, and A Talwalkar. A system for massively parallel hyperparameter tuning. arxiv 2018. *arXiv preprint arXiv:1810.05934*.
- [19] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
- [20] Matthias Rottmann, Lukas Pape, Mark Petersen, Hans Ulmer, and Hanno Gottschalk. Prediction error meta classification in semantic segmentation: Detection via aggregated dispersion measures of softmax probabilities. In *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pages 3045–3050, 2018.

- [21] Tianshu Wen, Kookjin Lee, and Youngsoo Choi. Reduced-order modeling for parameterized pdes via implicit neural representations. *arXiv preprint arXiv:2311.16410*, 2023.
- [22] Zhenxun Zhuang, Mingrui Liu, Ashok Cutkosky, and Francesco Orabona. Understanding adamw through proximal methods and scale-freeness. *Transactions on machine learning research*, 2022.