

Multiscale Attention Wavelet Neural Operator for Capturing Steep Trajectories in Biochemical Systems

Jiayang Su¹, Junbo Ma^{2,3*}, Songyang Tong¹, Enze Xu⁴, Minghan Chen^{4†}

¹Guangxi Key Lab of Multi-Source Information Mining and Security, Guangxi Normal University, Guilin 541000, China

²School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China

³Lishui Institute of Hangzhou Dianzi University, Hangzhou Dianzi University, Hangzhou 310018, China

⁴Department of Computer Science, Wake Forest University, NC 27109, USA

sujiayang2000@gmail.com, nudt_mjb@outlook.com, songyangt60@gmail.com, xezpku@gmail.com, chenm@wfu.edu

Abstract

In biochemical modeling, some foundational systems can exhibit sudden and profound behavioral shifts, such as the cellular signaling pathway models, in which the physiological responses promptly react to environmental changes, resulting in steep changes in their dynamic model trajectories. These steep changes are one of the major challenges in biochemical modeling governed by nonlinear differential equations. One promising way to tackle this challenge is converting the input data from the time domain to the frequency domain through Fourier Neural Operators, which enhances the ability to analyze data periodicity and regularity. However, the effectiveness of these Fourier based methods diminishes in scenarios with complex abrupt switches. To address this limitation, an innovative Multiscale Attention Wavelet Neural Operator (MAWNO) method is proposed in this paper, which comprehensively combines the attention mechanism with the versatile wavelet transforms to effectively capture these abrupt switches. Specifically, the wavelet transform scrutinizes data across multiple scales to extract the characteristics of abrupt signals into wavelet coefficients, while the self-attention mechanism is adeptly introduced to enhance the wavelet coefficients in high-frequency signals that can better characterize the abrupt switches. Experimental results substantiate MAWNO's supremacy in terms of accuracy on three classical biochemical models featuring periodic and steep trajectories. <https://github.com/SUDERS/MAWNO>.

Introduction

Biochemical modeling serves as a vital tool for unraveling the characteristics of biological processes, providing insights into essential cellular mechanisms and behaviors. Intriguingly, some foundational systems are marked by steep and dramatic changes in their dynamic trajectories—a phenomenon that shapes critical events in biology. For instance, cellular signaling pathways can exhibit abrupt switches that drive cells from quiescence to rapid proliferation or initiate critical defense responses. Illustrating this phenomenon, the Mitosis-Promoting Factor (MPF) model (Novak and Tyson 1993) elucidates how the abrupt changes of proteins regulate the progression of the cell through different phases of

the cell cycle. In Fig. 1, the concentration of cyclins rises dramatically, causing an increase in the activity of cyclin-dependent kinases (CDKs). This surge in MPF activity triggers the cell's entry into mitosis, the intricate process of cell division. Similarly, the Brusselator model (Prigogine 1978) demonstrates the sharp transitions between chemical oscillations and quiescence, underscoring the importance of comprehending and representing such shifts in dynamic systems. A fundamental challenge in this endeavor revolves around solving differential equations, particularly ordinary differential equations (ODEs), which form the backbone of dynamic system analysis. Traditional approaches often resort to numerical techniques like Euler's method (Hildebrand 1987) and Runger-Kutta method (Forsythe et al. 1977), discretizing the ODE into iterative computations of difference equations to approximate continuous solutions. Although these methods have provided a solid foundation, they still face limitations in capturing the sudden and profound behavioral shifts that characterize certain biological systems.

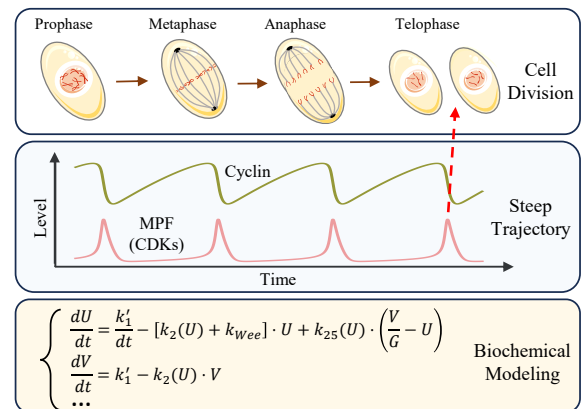


Figure 1: Illustration of Mitosis-Promoting Factor (MPF) model shows a cell undergoing sudden and pronounced changes in the trajectory of cyclin and MPF levels. This abrupt shift triggers the cell to enter mitosis (cell division).

*Corresponding author

†Corresponding author

In recent years, the advancement in machine learning techniques offers a refreshing perspective on solving differential equations, introducing innovative tools that adapt

to the complexities inherent in biochemical systems. At the forefront of this revolution are Physics-Informed Neural Networks (PINNs) (Raissi, Perdikaris, and Karniadakis 2019). By integrating physics knowledge into neural networks, PINN has become a data-efficient approach to solving ODEs of complex systems and has been sought after since its introduction due to its straightforward theory and easy-to-replicate architecture. Complementing this, DeepONet (Lu et al. 2021a) employs neural operators based on the generalized approximation theorem, presenting a data-driven strategy that excels in accuracy and flexibility. With its deeper network layers and adaptable architecture, DeepONet displays reduced errors and a capacity to tailor network structures for specific problem characteristics.

One notable advancement is the Fourier Neural Operator (FNO) (Li, Kovachki, and Azizzadenesheli 2021). FNO integrates neural networks with the Fourier transform to project input data from the time domain to the frequency domain. By parameterizing the kernel integral operator within the Fourier space, this transformation facilitates the extraction of analytical features and the approximation of nonlinear equations. Another subsequent work is wavelet neural operator (WNO) (Tripura and Chakraborty 2023). WNO leverages discrete wavelet transform (DWT) to efficiently decompose signals, thereby reducing the redundancy and computational overhead associated with continuous wavelet transform (CWT) techniques.

The effectiveness of FNO decreases when confronted with complex scenarios involving abrupt signals with steep changes in trajectories. This is because the Fourier transform, which utilizes the sine-cosine function as a basis, is good at mapping one-dimensional, triangular-wave-like continuous-variable function signals onto a sequence of coefficients, but is virtually impotent in the case of abrupt changes or high-dimensional, non-triangular-wave signals. In contrast, the wavelet transform projects signals into the wavelet domain, capturing both frequency and spatial information without smoothing the signal in the frequency domain as in the case of the Fast Fourier Transform (FFT). Therefore, WNO is capable of handling steep trajectories compared to neural operators such as FNO. However, it's important to note that as scenarios become more intricate, with a greater number of mutated signals, WNO's effectiveness also experiences a decline.

To enhance the neural network's ability to approximate equations featuring steep changes, a novel method called Multiscale Attention Wavelet Neural Operator (MAWNO) is proposed, which comprehensively combines wavelet transform with the self-attention mechanism in multiple wavelet scales. Specifically, the input is decomposed on multiple wavelet scales to obtain the high-level detail coefficients and approximate coefficients, which correspond to the signal's high-frequency and low-frequency information. Given that mutated signals predominantly manifest as high-frequency variations, the self-attention mechanism is strategically applied to the detail coefficients to capture distant dependencies in sequences. Furthermore, the proposed MAWNO also adopts the physics-driven approach, utilizing ODEs and their initial conditions to make predictions without

ground truth data labels. Experimental results substantiate its supremacy in accuracy when applied to three classical models featuring periodic and steep trajectories, surpassing PINN, FNO, and WNO methods. Overall, the salient points of the proposed MAWNO method are as follows:

- **Multiscale Self-Attentive Wavelet Design:** We introduce a distinctive wavelet block augmented by multiscale self-attention to enhance the mutual information from multiple scales of wavelet decomposition. This can solve nonlinear differential equations characterized by steep changes, a challenge that both PINN and FNO failed to address.
- **Efficient Self-Attention Implementation:** Self-attention is strategically applied to the coefficients after wavelet decomposition rather than to the entire raw signal, reducing the computational complexity associated with the attention mechanism.
- **Physics-Driven Approach:** MAWNO relies on physical/biochemical equations to establish an input-to-output mapping without ground truth data labels. This not only reduces the reliance on real-world data for model training but also augments the applicability of MAWNO across a broader spectrum of scenarios.

Related Works

Physics-Informed Neural Networks. PINNs (Raissi, Perdikaris, and Karniadakis 2019) is a learning framework to train neural networks with the constraints of physical laws described by differential equations. Compared with pure data-driven neural networks, PINN imposes physical information as constraints during the training process, so that a more general model can be learned with fewer data samples. Physics-Informed Neural Networks with Hard Constraints (hPINNs) (Lu et al. 2021b) achieved good results on reverse design problems by using penalty and augmented Lagrangian methods to impose hard constraints. To enhance the accuracy of PINNs further, Gradient-enhanced Physics-Informed Neural Networks (gPINN) (Yu et al. 2022) include the gradient information of the partial differential equations (PDEs) residual into the loss function. Applications of PINNs extend to diverse fields, including geology (Zheng, Zeng, and Karniadakis 2020), medicine (Raymond et al. 2022), fluid dynamics (Chen et al. 2022), and materials science (Chen et al. 2020), demonstrating their versatility and utility.

Neural Operators. Neural operators present a potent approach for solving differential equations through neural network-based methods, which entails learning nonlinear mappings between input and output function spaces. By harnessing neural operators, we can capture patterns and features of input data, and make accurate predictions on new and unseen data, thereby constructing versatile models. Neural operators were first introduced in the work of DeepONet (Lu et al. 2021a), which has been proven to be capable of approximating diverse nonlinear operators with a sufficiently large network. Building upon DeepONet, physics-informed DeepONets (Wang, Wang, and Perdikaris 2021)

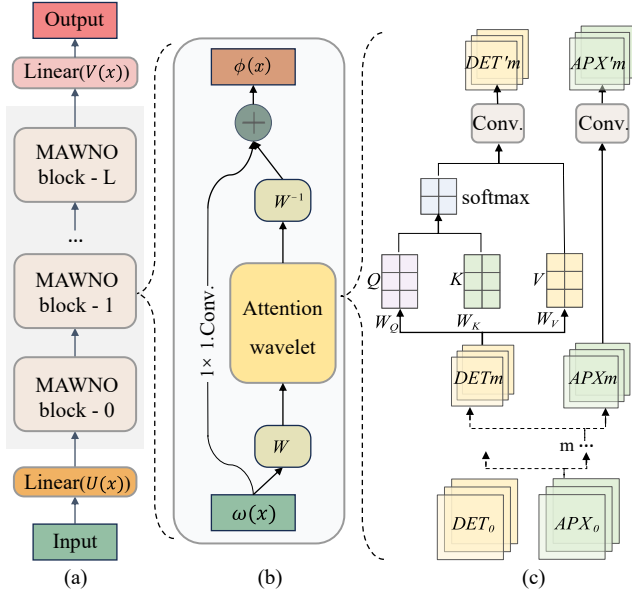


Figure 2: (a) MAWNO structure overview: MAWNO is mainly built from L MAWNO blocks. The $U(x)$ and $V(x)$ are two fully connected layers that help maintain the signal dimensions. (b) MAWNO block decomposition: the input $\omega(x)$ is firstly decomposed by wavelet transform W , then passed to the attention wavelet layer. After that, the inverse transform W^{-1} is applied to reconstruct the coefficients back to the feature domain. Meanwhile, a 1×1 convolution operation is performed on $\omega(x)$, and the results from two operations are combined to yield $\phi(x)$. (c) Wavelet attention layer of the m -th level: After the signal is decomposed into multiple levels, APX_m (m -level approximation coefficient) and DET_m (m -level detail coefficient) are derived. DET_m^* is obtained after applying self-attention to DET_m , and then the two coefficients of APX_m and DET_m^* are convolved to obtain APX_m' and DET_m' , respectively. Finally, APX_m' and DET_m' are reconstructed by wavelet inverse transform.

offers a simpler but more intuitive architecture with accelerated training capabilities. FNO (Li, Kovachki, and Aziz-zadenesheli 2021) learns the operator in the spectral domain using the Fast Fourier Transform (FFT), achieving a good balance between computational efficiency and analytical accuracy. Another contribution of FNO is that it can perform zero-shot super-resolution, that is, train at a lower resolution and evaluate at a higher resolution. As variants of FNO, MultiWavelet Transform (MWT) (Gupta, Xiao, and Bogdan 2021) and Wavelet Neural Operator (WNO) (Tripura and Chakraborty 2023) utilize the wavelet transform to process signals. WNO outperforms FNO and MWT on complex PDEs.

Beyond differential equations, neural operators find applications in diverse domains. Adaptive Fourier Neural Operators (AFNO) (Guibas et al. 2022) extends neural operators to the image domain.

Method

Problem Formulation

In this work, we consider ODEs in the bounded domain D , e.g., $[0, T]$, $D \in \mathbb{R}^d$ and the function space \mathcal{H} over D , including boundary shapes and source functions. And our goal is to learn an operator \mathcal{N} from the input function space \mathcal{A} to the solution space \mathcal{H} , i.e., $\mathcal{N} : \mathcal{A} \rightarrow \mathcal{H}$. The \mathcal{A} space can contain boundary conditions, initial points, differential equations, etc. We define an operator \mathcal{N} that maps \mathcal{A} space functions to \mathcal{H} space. Assuming $u \in \mathcal{A}$, $s \in \mathcal{H}$, the operator \mathcal{N} can be expressed as $\mathcal{N}[u_i; \lambda] = s_i, i \in [0, N]$, where $\lambda \in \Theta$ is the parameter space for the neural network to learn the \mathcal{N} operator mapping $\mathcal{A} \rightarrow \mathcal{H}$ and D is a subset of \mathbb{R}^d .

Our goal is to learn the operator by minimizing the loss function that combines the known initial value $u(0)$ and the ODE formula: $Loss = Loss_{u(0)} + Loss_f$. Here $Loss_{u(0)}$ aims to anchor the output to the initial condition of the ODE, while the $Loss_f$ is a residual loss between the derivative form of the ODEs and the neural network's outputs with corresponding derivatives through automatic differentiation. Minimizing $Loss_f$ will force the neural network to obey the constraints of the ODEs without additional data labels.

MAWNO Structure

The overall structure of MAWNO is shown in Fig. 2. The inputs are first projected into an embedding space through a fully connected layer and then fed into several MAWNO blocks. The MAWNO block is the basic built block of the proposed method, see Fig. 2 (b). The input of the MAWNO block is decomposed into wavelet coefficients by wavelet transform and then passed to the Attention wavelet layer for further processing. Also, the input of MAWNO is subjected to a convolution operation with a convolution kernel of 1×1 size. Since wavelet transform is an effective time-frequency analysis method, the time and frequency features can be effectively extracted. The 1×1 convolution kernel aims to linearly combine the features of different channels, thus achieving cross-channel feature fusion and adding non-linear features, which helps learn more complex feature representations. As shown in Fig. 2 (c), the input is decomposed into m levels by wavelet transform, where the lower level wavelet coefficients usually contain a lot of useless information such as noise, while the useful information is contained in the higher level wavelet coefficients. The m -th level wavelet coefficients consist of detail coefficients DET_m and approximation coefficients APX_m . The self-attention is strategically applied on DET_m , which allows the neural network to consider and correlate the ever-changing high-frequency information in the whole signal. DET_m performs self-attention to obtain DET_m^* , and the convolution operation is performed on DET_m^* and APX_m to analyze the inputs at multiple scales and extract the features of the inputs, in particular extracting the high-frequency information in DET_m^* . Thus enhancing the ability of MAWNO to approximate biochemical models with sharply varying trajectories.

Wavelet Transform. Wavelet transform is a time-frequency analysis method of signals. It has the character-

istics of multi-resolution analysis, and can characterize the local characteristics of signals in time-frequency domains. It can perform localized analysis of time (space) and frequency, and gradually perform multi-scale refinement of signals (functions) through stretching and translation operations. Therefore, it can focus on any details of the signal (Torrence and Compo 1998; Cotter 2019).

The wavelets represent sets of functions that result from dilation and translation from a single function, often termed as ‘mother function’, or ‘mother wavelet’. For a given mother wavelet $\psi(x)$, the resulting wavelets are written as,

$$\psi_{a,b}(x) = \frac{1}{|a|^{1/2}} \psi\left(\frac{x-b}{a}\right), \quad (1)$$

$$a, b \in \mathbb{R}, a \neq 0, x \in D$$

where a, b are the dilation, translation factor, respectively, and D is the domain of the wavelets under consideration. Let W and W^{-1} be the forward and the inverse wavelet transform of an arbitrary function $\Gamma : D \rightarrow \mathbb{R}^d$. Then, $(\Gamma_w)_j(a, b) = (W\Gamma)_j(a, b)\psi((x-b)/a) \in \mathcal{L}^2(\mathbb{R})$ is called the daughter wavelet, which is obtained by shifting and scaling of the mother wavelet $\psi(x)$.

Wavelet neural operator. As shown in Fig. 2 (b), the input data $u(t)$ passes through the fully connected layer $U(x)$ to map the features to a higher-dimensional space. This process aids the model in capturing complex patterns and relationships in the data, resulting in $v_0(t)$. The $v_0(t)$ serves as the input of the next layer, the MAWNO block layer. After passing through the L MAWNO blocks, the final output $v_L(t)$ is obtained, as $v_1 \mapsto \dots \mapsto v_j \mapsto \dots \mapsto v_L$. The data dimension is then reduced back to its original shape through the fully connected layer $V(x)$. The neural operator $\mathcal{N}(\cdot)$ can be formulated as,

$$\mathcal{N} := V \circ \sigma(G_L v_L(t) + \mathcal{K}_L + B_L) \circ \dots \circ \sigma(G_1 v_1(t) + \mathcal{K}_1 + B_1) \circ U \quad (2)$$

where σ is a fixed nonlinear activation function enhancing the nonlinear expression ability of the neural network, G is a linear transformation, B is the deviation of the neural network, and \mathcal{K} is the kernel integral operator. By parameterizing \mathcal{K} in the neural network to learn the neural operator \mathcal{N} . The kernel integral operator \mathcal{K} is defined as follows,

$$(\mathcal{K}(u) * v_j)(t) := \int_D k(u(t))v_j(t)dt, \quad (3)$$

$$t \in D, \quad j \in [1, L]$$

The convolution theorem, as articulated by Bracewell in 1966 (Bracewell and Kahn 1966), establishes that the act of convolving two signals in the time domain is equivalent to multiplying their respective Fourier transforms in the frequency domain. Conversely, convolving two signals in the frequency domain is tantamount to the multiplication of their inverse Fourier transforms in the time domain. Consequently, following the application of the convolution theorem to our kernel integral operator \mathcal{K} , we rewrite Eq. (3) as a wavelet neural operator:

$$(\mathcal{K}(u) * v_j)(t) = W^{-1}(W(k_\phi) \cdot W(v_j))(t), t \in D. \quad (4)$$

where W and W^{-1} denote the wavelet transform and wavelet inverse transform, respectively, and we omit the subscripts \mathcal{K}, W to simplify the equation.

Multiscale Wavelet Attention

The standard self-attention mechanism is applied to the detail coefficients at the highest level of wavelet decomposition. Suppose there are 3 sequences Q, K, V , which are defined as follows,

$$\begin{aligned} Q &= DET_m * W_Q, \\ K &= DET_m * W_K, \\ V &= DET_m * W_V. \end{aligned} \quad (5)$$

where W_Q, W_K, W_V are learnable parameter matrices, defining the attention formula,

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

Combining Eqs. (5) and (6), DET'_m can be easily obtained:

$$DET'_m = \text{Attention}(Q, K, V) \quad (7)$$

The main idea of Self-attention (SA) is to allow the model to automatically learn the association and importance between different positions based on the information at different positions in the input sequence (Vaswani et al. 2017). Addressing the shortcomings of traditional Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) that usually only focus on local information, SA is able to take into account the information of the whole sequence at the same time, thus better capturing long-distance dependencies in the sequence. Combining the attention mechanism with other methods is a popular research, such as AEC-LSTM (Huang et al. 2021). Our proposed MAWNO innovatively applies self-attention to wavelet transformation. The input signal is decomposed into m levels through discrete wavelet transform (DWT). Each decomposition will produce a pair of approximate coefficients and detail coefficients. The detail coefficients contain high-frequency features, and the approximate coefficients contain low-frequency features. Usually, the lower-level high-frequency signal is composed of noise, and the higher-level wavelet coefficients contain the most critical information of the entire signal. Applying the self-attention mechanism to the detail coefficient DET_m after wavelet decomposition m level in the wavelet domain can capture the correlation between the detail coefficients, and enhance the neural network’s ability to approximate equations featuring steep changes.

Experiments and Results

Experimental Settings

The effectiveness of the proposed MAWNO is validated through a series of experiments and compared with three state-of-the-art methods, namely, PINN, FNO, and WNO:

- **PINN** (Raissi, Perdikaris, and Karniadakis 2019): Physics-Informed Neural Network uses neural networks to solve mathematical equations infused with physical principles. This method is valued for its simplicity and efficiency.

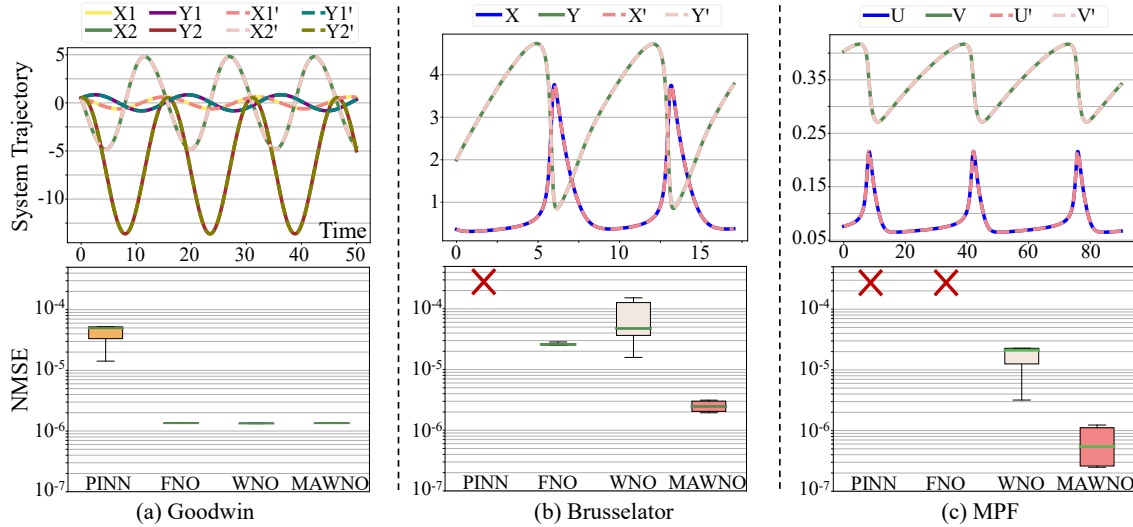


Figure 3: Top row: ground truth (LSODA solver, solid lines) and predicted trajectories (MAWNO, dashed lines) of three biochemical models, including (a) Goodwin, (b) Brusselator, and (c) MPF. Bottom row: the box plot of NMSEs obtained by PINN, FNO, WNO, and MAWNO, across five independent experiments. The red cross sign indicates cases of failure that deviate significantly from the ground truth.

- **FNO** (Li, Kovachki, and Aizzadenesheli 2021): Fourier Neural Operator processes signals in spectral space, rendering its learning approach effective.
- **WNO** (Tripura and Chakraborty 2023): Wavelet Neural Operator, an extension of FNO, projects the signal into the wavelet domain and parameterizes the wavelet neural operator to solve PDEs.

MAWNO uses sym9, sym9, and db6 as wavelet bases for Goodwin, Brusselator, and MPF, respectively. Wavelet basis is the basic function used to decompose and reconstruct signals. Different types of wavelet bases have different frequency and scale characteristics, which can be used to capture different frequency components and characteristics of signals. We choose Daubechies family (Daubechies 1992) and Symlet family (Meyer 1993) wavelets as wavelet base functions of MAWNO. The number of wavelet decomposition layers m is 8 for all three biochemical models. The wavelet level indicates the number of input wavelet decomposition layers. The higher the number of layers, the stronger the ability to capture high-frequency signals, but it will also increase the computational complexity, and an excessively high number of layers will increase the risk of model overfitting. The number of MAWNO blocks is set to 4. MAWNO block depth represents the number of MAWNO blocks in the model. If the number is too small, it will lead to underfitting, and if the number of layers is too deep, it will increase the difficulty of model training. All models are trained with the Adam (Kingma and Ba 2015) optimizer that adaptively optimizes the learning rate. We uniformly sampled 2^{13} sample points for Goodwin, Brusselator, and MPF, and trained them for 50000, 200000, and 500000 epochs, respectively.

The Normalized Mean Square Error (NMSE) is utilized as the evaluation metric. NMSE is a normalized form of mean

square error, which normalizes the MSE value to a certain range, making it easier to compare the performance of different systems. The relative error of NMSE is calculated as:

$$\text{NMSE}(\%) = \frac{1}{N} \sum_{i=1}^N \frac{\|u'_i - u_i\|_2}{\|u_i\|_2} \times 100\% \quad (8)$$

where $\|\cdot\|_2$ is the 2-norm, N is the total number of samples, and u'_i is the prediction of the ground truth u_i .

All four methods are evaluated on three classical biochemical models featuring periodic and steep trajectories, including Goodwin, M-phase Promoting Factor (MPF), and Brusselator.

Trajectory Prediction of Biochemical Systems

Goodwin (Goodwin 1965): The Goodwin model elucidates the rhythmic behavior exhibited by various biological systems, including the circadian clock, cell cycle, and segmentation clock. The model comprises three equations delineating the dynamics of a gene, its corresponding mRNA, and the resultant protein. An inhibitory relationship is established wherein the protein suppresses gene expression, manifesting a negative feedback loop. Renowned for its simplicity and widespread adoption, the Goodwin model is a cornerstone in the realm of biological oscillations. The corresponding ordinary differential equations are given below.

$$\begin{cases} \frac{dX_1}{dt} = \frac{a_1}{A_1 + k_{11}Y_1 + k_{12}Y_2} - b_1 \\ \frac{dY_1}{dt} = a_1X_1 - \beta_1 \\ \frac{dX_2}{dt} = \frac{a_2}{A_2 + k_{21}Y_1 + k_{22}Y_2} - b_2 \\ \frac{dY_2}{dt} = a_2X_2 - \beta_2 \end{cases}$$

The experimental results are shown in Fig. 3 (a). The trajectories predicted by our MAWNO align well with the ground truth obtained using the ODE solver (LSODA). It can be seen that FNO, WNO, and MAWNO perform consistently on the Goodwin system, yielding remarkably low NMSE at approximately 1.3×10^{-6} . In contrast, PINN’s accuracy ranges from $1 \times 10^{-5} \sim 5 \times 10^{-5}$, and its variances fall short in comparison to the other three models.

Brusselator (Prigogine 1978): The Brusselator is designed to study auto fluctuations and spatial structure formation in nonequilibrium systems. The model can form sharp oscillatory patterns marked by abrupt transitions between high and low concentrations in a periodic cycle. Its relevance extends to diverse natural phenomena like biochemical reactions, nonlinear optics, and electrochemical reactions, contributing to our understanding of many non-equilibrium systems in nature. In our experiments, we use the Brusselator constant differential formulation from (Lozano-Parada, Burnham, and Machuca Martinez 2018), given as follows.

$$\begin{cases} \frac{d(X)}{dt} = A - B \cdot X + X^2 \cdot Y - X \\ \frac{d(Y)}{dt} = B \cdot X - X^2 \cdot Y \end{cases}$$

In Fig. 3 (b), PINN fails to predict the Brusselator model, underscoring its ineffectiveness for systems characterized by high complexity and sharpness. Among the evaluated methods, MAWNO achieves the smallest NMSE values within the range of $2 \times 10^{-6} \sim 4 \times 10^{-6}$. Following MAWNO, FNO shows a slightly higher NMSE of approximately 2×10^{-5} , albeit with reduced variance. WNO’s performance is less robust compared to MAWNO and FNO, yielding an NMSE range of $1 \times 10^{-5} \sim 1 \times 10^{-4}$.

MPF (Novak and Tyson 1993): M-phase Promoting Factor is a heterodimer composed of the Cdc2 protein and cell cycle proteins. These MPF complexes play important roles in regulating cell division at different stages of the cell cycle. The MPF model investigates the interactions between cdc2 and cyclin to reveal their dynamic behaviors and regulatory mechanisms in the cell cycle. These studies are of great significance for the in-depth understanding of cell cycle regulation, the precise control of cell division, and the fundamental mechanisms of cell biology. The dynamic system of MPF is described below.

$$\begin{cases} \frac{dU}{dt} = \frac{k'_1}{dt} - [k_2(U) + k_{Wee}] \cdot U + k_{25}(U) \cdot \left(\frac{V}{G} - U\right) \\ \frac{dV}{dt} = k'_1 - k_2(U) \cdot V \\ k_2(U) = k'_2 + k''_2 U^2 \\ k_{25}(U) = k'_{25} + k''_{25} U^2 \end{cases}$$

In Fig. 3 (c), the MPF system showcases remarkably steep trajectories culminating in a sharp apex. MAWNO distinctly stands out, as its predicted paths align consistently with the ground truth. This exceptional alignment illustrates MAWNO’s capacity to adeptly manage systems with substantially higher complexity and sharpness. In contrast, neither PINN nor FNO proves effective in this case. While

WNO achieves an NMSE span of $2 \times 10^{-5} \sim 5 \times 10^{-5}$, MAWNO surpasses this with an NMSE ranging from $2 \times 10^{-7} \sim 1 \times 10^{-6}$, one order of magnitude improvement.

Performance Analysis on Varying Steepness

To validate the effectiveness of MAWNO in handling equations characterized by steep trajectories, we carried out a comparative assessment involving PINN, FNO, WNO, and MAWNO. This evaluation was performed on the Brusselator model by systematically changing the parameters to vary the degree of the trajectory’s steepness, as depicted in Fig. 4. Evidently, PINN fails in all four cases of different steepness levels. In Fig. 4 (a) and 4 (b), where the Brusselator system displays smooth and moderate dynamics, the accuracy of MAWNO is akin to those of FNO and WNO. As the trajectory becomes steeper, the advantages of MAWNO become prominent. As shown in Fig. 4 (c) and (d), MAWNO outperforms FNO and WNO by a significant margin. The culmination of these four experiments with increasing steepness underscores MAWNO’s superior aptitude in effectively addressing dynamical systems characterized by abrupt changes in trajectories.

Parameter Sensitivity Experiments

The parameter sensitivity experiments are done on the Brusselator model. Fig. 5 (a-c) shows the effect of different wavelet decomposition layers m , different wavelet bases, and dataset sizes in terms of NMSE and training duration. As n increases in wavelet basis sym n , the length of wavelet basis support moments increases. The experimental results show that larger decomposition layers m , longer support moments, and larger dataset size increase the training time length. So combining the accuracy and time consumption, the suitable parameters are chosen.

Additionally, we investigate the impact of varying the number of layers within the MAWNO block, see Fig. 5 (d). Employing a single layer or two layers for the MPF model yielded unsuccessful results. It can be seen that as the number of layers increases, the NMSE decreases until a nadir. After reaching the fourth layer, the NMSE starts to rise. This is likely due to the network’s depth becoming excessively profound, thereby contributing to the model’s degradation.

Discussion

The experimental results demonstrate MAWNO’s superiority as a powerful and versatile approach. Its impressive adaptability and the lowest NMSE error among the four methods establish MAWNO as the top choice for a diverse spectrum of biochemical systems, from simple to complex, setting new standards for precision and efficiency in solving differential equations in biochemical modeling.

The limitations of PINN stem from its simplistic neural network architecture, rendering it less effective in addressing the intricacies of complex systems. FNO encounters difficulties when confronted with steep trajectories, which hinders its overall performance. WNO exhibits advantages over both FNO and PINN in navigating abrupt dynamic changes, allowing it to excel in the challenging MPF model where

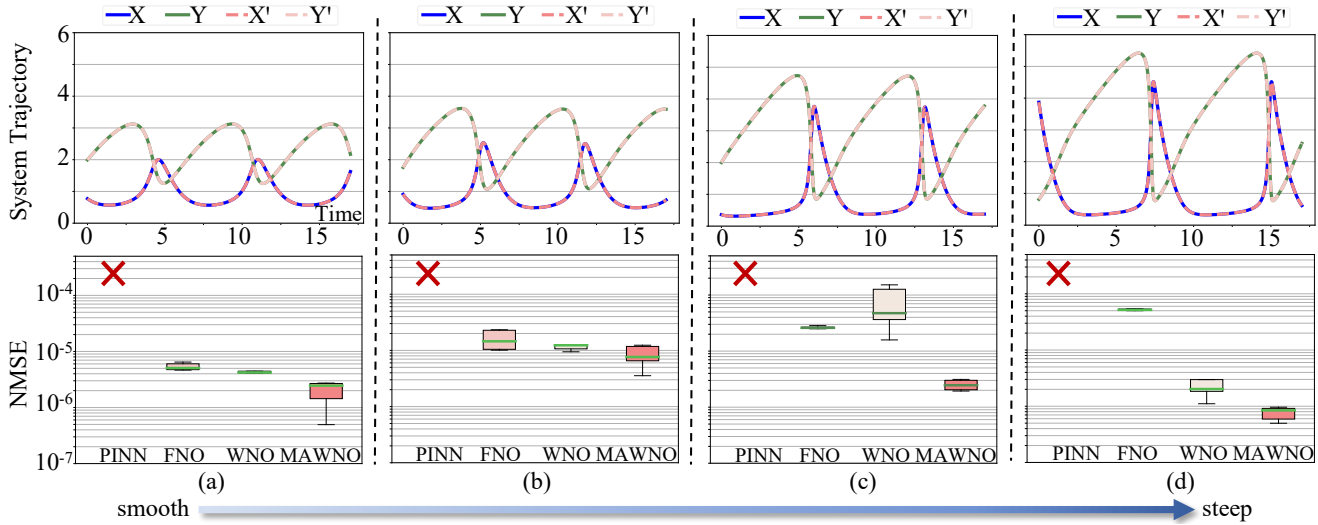


Figure 4: Top row: the real situation (LSODA solver, solid lines) and predicted trajectory (MAWNO, dotted lines) of four Brusselators with increasing steepness from (a) to (d), obtained by varying the parameters of the original Brusselator. Bottom row: the box plot of NMSEs obtained by PINN, FNO, WNO, and MAWNO, across five independent experiments.

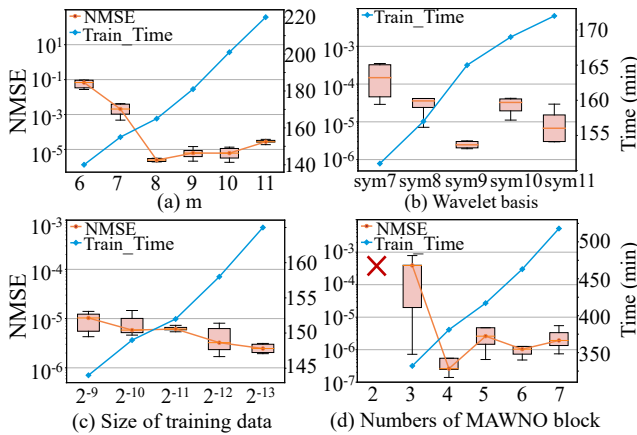


Figure 5: Parameter sensitivity experiments (a-c) on the Brusselator model: NMSE and training time (blue line) with varying numbers of wavelet decomposition layer m (a), types of wavelet basis (b), and sizes of training data (c). Experiment (d) shows the NMSE values and training time of the MPF model under different numbers of MAWNO blocks. The training time remains consistent across different random seeds.

FNO and PINN fail. However, it is crucial to note that WNO comes with a trade-off, as it carries a higher NMSE compared to the more precise MAWNO. Furthermore, when dealing with moderately complex models like the Brusselator, MAWNO proves to be more effective than WNO. This highlights MAWNO’s versatility across a wide range of scenarios as well as higher precision when compared to WNO. MAWNO prevails as a robust solution capable of handling not only simple and smooth models, but also sharp and com-

plex ones with remarkable accuracy.

Conclusion

In this paper, a novel Multiscale Attention Wavelet Neural Operator (MAWNO) method is proposed to tackle the challenge of abrupt switches in biochemical systems described by nonlinear differential equations. The proposed MAWNO comprehensively combines the self-attention mechanism with the versatile wavelet transforms to enhance the multiscale coefficients obtained from multiple levels of wavelet decompositions, capturing the fine-grained and distant information in the input signals. Moreover, MAWNO is also a physics-driven approach that can efficiently exploit biochemical equations to establish an input-to-output mapping without ground truth data labels. Extensive experiments demonstrate the effectiveness of the proposed MAWNO on three representative biochemical models featuring periodic and steep trajectories, surpassing the state-of-the-art methods in terms of accuracy.

Acknowledgements

This work was supported in part by National Key Research and Development Program of China under Grant 2022YFA1004100, the Guangxi Natural Science Foundation (2023GXNSFBA026010), Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (22-A-03-02), and Innovation Project of Guangxi Graduate Education (XJCY2022009). We would like to thank the anonymous referee for valuable suggestions that helped us to significantly improve this paper.

References

Bracewell, R.; and Kahn, P. B. 1966. The Fourier transform and its applications. *American Journal of Physics*, 34(8):

712–712.

Chen, D.; Gao, X.; Xu, C.; Wang, S.; Chen, S.; Fang, J.; and Wang, Z. 2022. FlowDNN: a physics-informed deep neural network for fast and accurate flow prediction. *Frontiers of Information Technology & Electronic Engineering*, 23(2): 207–219.

Chen, Y.; Lu, L.; Karniadakis, G. E.; and Dal Negro, L. 2020. Physics-informed neural networks for inverse problems in nano-optics and metamaterials. *Optics express*, 28(8): 11618–11633.

Cotter, F. B. 2019. *Uses of Complex Wavelets in Deep Convolutional Neural Networks*. Ph.D. thesis, Department of Engineering, University of Cambridge, Cambridge, UK.

Daubechies, I. 1992. *Ten lectures on wavelets*. Society for industrial and applied mathematics.

Forsythe, G. E.; et al. 1977. *Computer methods for mathematical computations*. Prentice-hall.

Goodwin, B. C. 1965. Oscillatory behavior in enzymatic control processes. *Advances in enzyme regulation*, 3: 425–437.

Guibas, J.; Mardani, M.; Li, Z.; Tao, A.; Anandkumar, A.; and Catanzaro, B. 2022. Efficient Token Mixing for Transformers via Adaptive Fourier Neural Operators. In *International Conference on Learning Representations*.

Gupta, G.; Xiao, X.; and Bogdan, P. 2021. Multiwavelet-based operator learning for differential equations. *Advances in neural information processing systems*, 34: 24048–24062.

Hildebrand, F. B. 1987. *Introduction to numerical analysis*. Courier Corporation.

Huang, F.; Li, X.; Yuan, C.; Zhang, S.; Zhang, J.; and Qiao, S. 2021. Attention-emotion-enhanced convolutional LSTM for sentiment analysis. *IEEE transactions on neural networks and learning systems*, 33(9): 4332–4345.

Kingma, D. P.; and Ba, J. L. 2015. Adam: A method for stochastic gradient descent. In *ICLR: international conference on learning representations*, 1–15. ICLR US.

Li, Z.; Kovachki, N. B.; and Azizzadenesheli, K. 2021. Burigede liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*.

Lozano-Parada, J. H.; Burnham, H.; and Machuca Martinez, F. 2018. Pedagogical approach to the modeling and simulation of oscillating chemical systems with modern software: the Brusselator model. *Journal of Chemical Education*, 95(5): 758–766.

Lu, L.; Jin, P.; Pang, G.; Zhang, Z.; and Karniadakis, G. E. 2021a. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3): 218–229.

Lu, L.; Pestourie, R.; Yao, W.; Wang, Z.; Verdugo, F.; and Johnson, S. G. 2021b. Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43(6): B1105–B1132.

Meyer, Y. 1993. *Wavelets: algorithms & applications*. Philadelphia: SIAM (Society for Industrial and Applied Mathematics).

Novak, B.; and Tyson, J. J. 1993. Modeling the cell division cycle: M-phase trigger, oscillations, and size control. *Journal of theoretical biology*, 165(1): 101–134.

Prigogine, I. 1978. Time, structure, and fluctuations. *Science*, 201(4358): 777–785.

Raissi, M.; Perdikaris, P.; and Karniadakis, G. E. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378: 686–707.

Raymond, S. J.; Cecchi, N. J.; Alizadeh, H. V.; Callan, A. A.; Rice, E.; Liu, Y.; Zhou, Z.; Zeineh, M.; and Camarillo, D. B. 2022. Physics-informed machine learning improves detection of head impacts. *Annals of biomedical engineering*, 50(11): 1534–1545.

Torrence, C.; and Compo, G. P. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1): 61–78.

Tripura, T.; and Chakraborty, S. 2023. Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. *Computer Methods in Applied Mechanics and Engineering*, 404: 115783.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, S.; Wang, H.; and Perdikaris, P. 2021. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Science advances*, 7(40): eabi8605.

Yu, J.; Lu, L.; Meng, X.; and Karniadakis, G. E. 2022. Gradient-enhanced physics-informed neural networks for forward and inverse PDE problems. *Computer Methods in Applied Mechanics and Engineering*, 393: 114823.

Zheng, Q.; Zeng, L.; and Karniadakis, G. E. 2020. Physics-informed semantic inpainting: Application to geostatistical modeling. *Journal of Computational Physics*, 419: 109676.