Research on Prediction of Lorenz System Based on Combined Physical Model with Machine Learning Model

Student Name: Jiaxuan Sun

Supervisor Name: Hailiang Du

Submitted as part of the degree of M.Sc. MISCADA to the

Board of Examiners in the Department of Computer Science, Durham University.

Abstract – This study investigates the synergistic combination of machine learning model (Gradient Boosting Decision Trees (GBDT) and Echo State Networks (ESN) with imperfect physical models for predicting the chaotic Lorenz system. To that end, we constructed a parametrically controllable imperfect physical model and generated simulated observational sequences with added noise. Preliminary experiments indicated superior predictive accuracy of the ESN model over GBDT when relying solely on observational data. Furthermore, we tested two distinct methods of integrating the physical model. One approach comprised concurrent input of the physical model's prediction values alongside observed data, while the other entailed learning solely the residual between the physical model and observations. The results demonstrated that the former approach significantly enhanced the predictive accuracy of both models, whereas the latter was less efficacious. Notably, the ESN model persistently outperformed the GBDT model in terms of predictive capability. This study provides a beneficial exemplar of the hybrid modelling of machine learning and physical models, establishing a foundation for further enhancing the prediction of complex dynamical systems.

Keywords – Nonlinear dynamical systems, Lorentz systems, Gradient Boosting Decision Trees, Echo State Networks.

I. INTRODUCTION

Dynamical Systems, as defined by Connell et al. (2017), constitute a theoretical framework used to comprehend and predict self-organizing phenomena within complex systems, which evolve, reconfigure, and develop over time. Dynamical Systems are typically employed to encapsulate the processes of change within specific systems. Since its inception, this framework has found widespread application across various disciplines, such as meteorological patterns (Wiggins, 2005), economic markets (Brock & Dechert, 1991), traffic flow (Chowdhury et al., 2000), and ecosystems (Berryman, & Millstein, 1989). Differential equations have emerged as a focal point in describing dynamical systems, given their intricate interplay of multiple influences, complex behavior, and nonlinearity (Zhao et al., 2023). The intricate interdependencies among multiple factors often lead to intricate nonlinear behaviors in these systems, sometimes even giving rise to chaotic dynamics (Kloeden & Mees, 1985). Accurate modeling and long-term prediction of these systems pose substantial challenges due to the complex nonlinear relationships governing their behavior (Peter A. G. Watson, 2019).

Machine learning has become a solution to address predictive challenges in intricate dynamic systems (Bradley & Kantz, 2015). This approach entails employing observed data from dynamic systems and utilizing statistical or machine learning models to fit the acquired data.

Machine learning models often make weak or no assumptions about the linear relationship between data features and data distribution. As a result, they possess better generalization capability and broader applicability, thereby facilitating the prediction of the system's future states or an in-depth

analysis of its features. The application of these models has yielded success in various industries, including finance (Wen et al., 2019) and weather forecasting (Hewage et al., 2020).

However, the sensitivity of chaotic systems to small errors makes forecasts unreliable and poses a significant obstacle to long-term predictions. In addition, complex nonlinear relationships make model fitting and generalization difficult, further leading to prediction failure for machine learning models. Traditional physics-based models rely on physical laws, but it is difficult to consider all complex physical relationships in the modeling process, or we need to make certain simplifying assumptions, so the established physical models are not perfect (Xu et al. , 2022), and in order to obtain accurate model results, it is often necessary to use high-precision numerical solution methods, but the use of high-precision numerical solution methods may bring huge computational pressure (Fu et al., 2023). Therefore, for predicting nonlinear dynamical systems, both traditional physical models and machine learning models have their own shortcomings.

To surmount these challenges, this paper explores the amalgamation of machine learning models and physics-based models. We investigate how machine learning models can rectify the imperfections in physics-based models, aiming to achieve more accurate and dependable predictions. Through this integration, we aspire to identify which information from physical models can enhance the performance of machine learning and thus enhance our ability to predict the behavior of dynamical systems.

The Lorenz system, renowned for its rich dynamical behavior (Algaba et al., 2013) and prediction difficulties, has been extensively discussed in hundreds of papers (Llibre et al., 2010), making it a classic subject of nonlinear dynamical system research. Therefore, this paper also employs the Lorenz 63 system as the target system to investigate the effectiveness of combining machine learning models with physics-based models.

The findings of this study hold theoretical and practical significance in enhancing the accuracy of dynamic system predictions. This not only aids in a better understanding and prediction of complex system behaviors but also provides more effective decision support and management strategies across diverse domains. Ultimately, the outcomes of this research are poised to contribute to the realization of more intelligent and dependable predictive models and decision systems.

II. LITERATURE REVIEW

Currently, numerous research efforts are devoted to constructing machine learning and deep learning models suitable for nonlinear time series analysis. Among these models, many exhibit outstanding performance. In this context, deep learning models stand out due to their rapid iteration speed. However, some studies have expressed reservations about the performance of deep learning models. Within the realm of machine learning models, a focal point lies in Echo State Network, which demonstrate promising performance in dealing with nonlinear time series. Nevertheless, does the progress of these machine learning models imply that we can discard physical models? Addressing this question, several research endeavors are dedicated to the integration of physical models and machine learning models. This section will separately delve into recent developments in deep learning models, the Echo State Network model, as well as the fusion of physical and machine learning models.

A. Machine Learning Models

There are numerous machine learning and deep learning models designed for time series analysis, and the evolution of models in this field is rapid. Among these, the most recent standout model is TimesNet, introduced by Wu et al. (2022). They contend that treating time series as low-dimensional observations alone makes long-term prediction challenging. Thus, from their investigation into time series, they propose a perspective wherein time series can be perceived as compositions of multiple cycles. Employing discrete Fourier transformation, they expand one-dimensional time series into two-dimensional tensors. Through the utilization of the Time Block method, they adaptively extract intricate temporal patterns and facilitate long-term forecasting.

Zhou et al. (2022) also recognize the difficulty of conventional time series models in capturing comprehensive trends and enabling extended-term predictions. To address this, they fuse the Transformer architecture with seasonal trend decomposition. Additionally, they note that time series data often lend themselves to sparse representations, such as Fourier transformation. They amalgamate this insight with seasonal decomposition to introduce the Frequency Enhanced Decomposed Transformer (FEDformer), thereby enhancing the performance of traditional Transformers by 14%. However, the structure of Transformer models, featuring encoder-decoder components, leads to high memory usage, and their quadratic time complexity presents training challenges. Thus, Zhou et al. (2021) innovate upon this architecture, introducing Informer. They incorporate a generative-style decoder, reducing its time complexity to $O(LlogL)$. Simultaneously, they revise its iterative methodology, substantially augmenting inference speed. The incorporation of the self-attention distilling highlights mechanism sharpens attentional focus, thereby enhancing its capability to handle lengthy time series data.

Although Transformer models have emerged as a focal point of research, resulting in a profusion of outstanding models, Zeng et al. (2023) propose an alternative perspective. Despite the benefits of employing positional encoding and token embedding for preserving some order information within Transformers, the inherent nature of permutation-invariant self-attention mechanisms inevitably leads to temporal information loss. Through a comparison of a straightforward single-layer network and the intricate Transformer across nine real-world datasets, their findings consistently demonstrate the superiority of the single-layer network, often by a substantial margin. Elsayed et al. (2021) similarly cast doubts on the contemporary application of deep learning in time series analysis. Through a comparative study involving Gradient Boosting Regression Trees (GBRT) and eight prominent models, they ascertain that the efficacy of simple GBRT surpasses that of deep learning networks. Consequently, the necessity of deploying deep learning techniques for time series analysis might warrant further investigation in future research endeavors.

B. Echo State Networks

In the realm of forecasting chaotic time series generated by complex systems, a focal point of research is the Echo State Network (ESN). Jaeger & Haas (2004) established its significant enhancement in predicting chaotic time series compared to conventional methods. Lukoševičius & Jaeger (2009) further highlighted ESN's superior predictive capabilities for nonlinear and chaotic systems in contrast to other network architectures.

Traditional feedforward neural networks are limited in their ability to learn the temporal relationships between inputs and outputs, thus exhibiting weaker capabilities in processing time series data. To overcome this issue, Recurrent Neural Networks (RNNs) were introduced and demonstrated effective learning of temporal correlations (Rumelhart, Hinton & Williams, 1986). However, due to the long-term dependencies within the hidden states of RNNs, the use of backpropagation through time (Werbos, 1990) involving repeated multiplicative gradient updates could lead to either vanishing or exploding gradients (Bengio et al., 1994). As a remedy, Jaeger & Haas (2004) introduced the concept of Echo State Networks (ESNs), a variant of RNNs.

Echo State Networks employ the notion of utilizing the neural network's hidden layer as the state space model for processing sequential data. ESNs achieve this by mapping input data into a higher-dimensional reservoir space, and subsequently employing an output matrix to project the learned states into the phase space, thereby facilitating temporal prediction and sequence processing. In contrast to conventional RNNs, ESNs require training solely for the output layer, employing a straightforward linear regression process. This approach mitigates the challenges posed by the repeated multiplication of gradients, thus circumventing issues related to gradient vanishing and explosion.

The architectural innovation of ESNs provides an elegant solution for addressing the difficulties encountered in RNNs while handling temporal data, reinforcing their efficacy in learning and predicting sequential patterns.

The structure of the Echo State Network is illustrated in Figure 1.

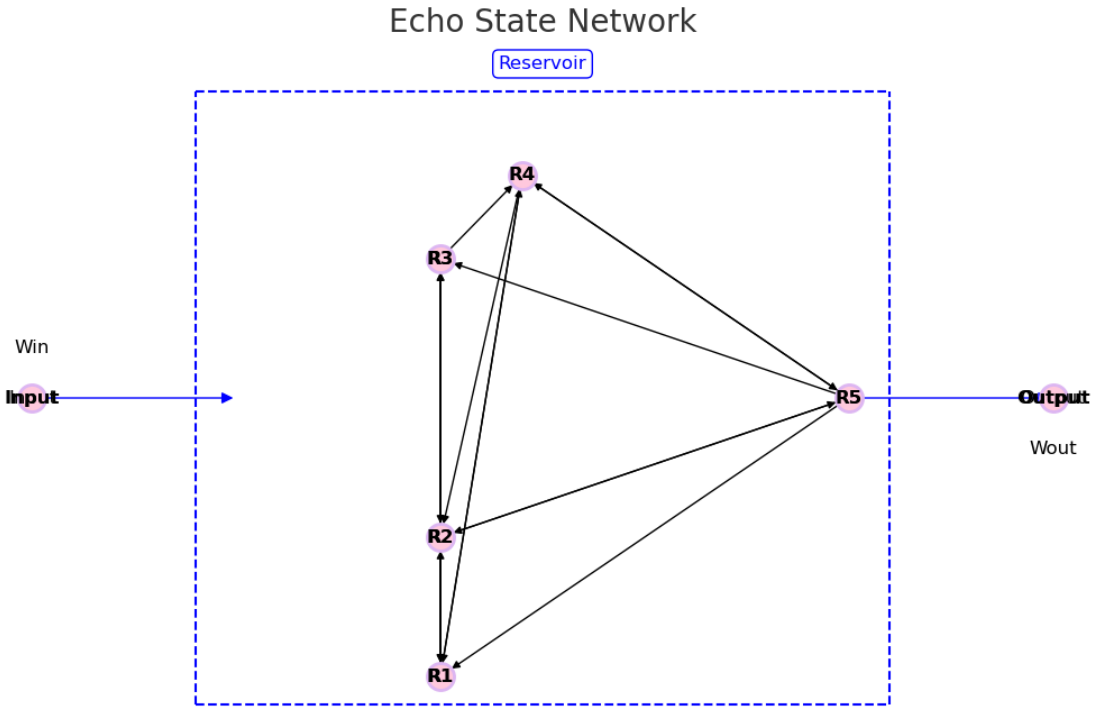

Figure 1. Network Architecture of the Echo State Network (ESN)

At time t, the input layer receives the input vector $\mathrm{u}(t)$, which, upon passing through input weights W_{in}, yields the input state $\mathrm{x_{in}(t)}$, where $\mathrm{x_{in}(t)} = \mathrm{W_{in}u(t)}$. Following processing at the input nodes, the data enters the network's second layer, referred to as the reservoir, among these, the reservoir comprises N nodes, denoted as $R1$ to RN. At time t, the matrix formed by concatenating the data stored in these nodes constitutes the state of the reservoir $x(t)$. The initial stage $\mathrm{x(0)}$ is a random vector. At each time t, the network state updates based on the recurrent weight matrix W, considering both the previous time step's state and input state. This update process can be formulated as follows:

$$\mathrm{x(t)} = \tanh\left(\mathrm{Wx(t-1)} + \mathrm{x_{in}\,(t)}\right)$$

where W is the recurrent weight matrix, representing the random connections between various states in the reservoir by multiplication with the previous time step's reservoir state. Initially, it is a matrix of randomly assigned weights, which is subsequently recalibrated across the entire matrix using a multiplication factor to set the spectral radius less than 1, because A spectral radius greater than 1 can lead to unstable behavior, while a spectral radius less than 1 generally results in stable dynamics (Yildiz et al., 2012). The final reservoir result is then processed through output weights W_{out} to yield the output:

$$\hat{\mathrm{y}}(\mathrm{t}) = \mathrm{W_{out}x(t)}$$

Within the network, only the output weights require training, while input and connection matrices are randomized. The training of the output matrix involves minimizing the Mean Squared Error (MSE) between the output $\hat{\mathrm{y}}(\mathrm{t})$ and the observation data $\mathrm{y(t)}$ on the training set on sample size N_{tr}.

$$\mathrm{MSE} \triangleq \frac{1}{N_{tr}} \sum_{t=1}^{N_{tr}} \|\mathrm{y(t)} - \hat{\mathrm{y}}(\mathrm{t})\|^2$$

where$\|\cdot\|$represents the L2 norm. The minimization process can be achieved using Ridge Regression, which is expressed as:

$$(\mathrm{x(t)x(t)}^T + \beta I)\mathrm{W}_{out} = \mathrm{x(t)y(t)}^T$$

$\mathrm{x(t)}$ represents the state of the reservoir, $\mathrm{y(t)}$ stands for observed data, I denotes the identity matrix, and β is user-defined Tikhonov regularization parameter (Tikhonov et al., 1995). Controlling the complexity of a model, prevent overfitting, and enhance the model's generalization ability.

The reservoir dimension of Echo State Networks is often significantly larger than the input dimension. The process of mapping input data to a high-dimensional reservoir through input weights corresponds to projecting data from low-dimensional to high-dimensional space. This approach has been substantiated by numerous studies as highly effective for processing nonlinear data (Balcan et al., 2006), endowing Echo State Networks with robust capabilities for handling nonlinear data.

C. Combining Physical and Machine Learning Models

Although machine learning models have shown strong capabilities, many studies have found that combining physical models can further improve the capabilities of machine learning models. The amalgamation of physical and machine learning models has demonstrated substantial enhancement in various domains. Employing machine learning models to predict residuals between physical models and observed data contributes to improved generalization and interpretability of the machine learning models. Furthermore, it can augment the predictive efficacy of physical models (Ren et al., 2022). Farchi et al. (2021) employed data assimilation techniques in conjunction with machine learning, introducing machine learning trend terms into physical models, resulting in a significant enhancement of predictive capabilities. Tu et al. (2023) integrated physical and machine learning models to simulate lithium-ion batteries. By providing the machine learning model with data from the physical model, profound physical insights were imparted, leading to substantial predictive accuracy improvements. Yang et al. (2020) adopted a similar approach, furnishing machine learning models with data amalgamating observations and physical models, thereby facilitating hydrological model development. The findings underscored pronounced performance gains arising from this integration.

In summary, while current deep learning methodologies have yielded certain achievements, the underlying escalation in computational costs necessitates further discussions regarding their model effectiveness. Conversely, simple methods such as Gradient Boosting Regression Trees (GBRT) and single-layer neural networks have exhibited promising performance. Moreover, Echo State Networks, renowned for their potent mapping capabilities from low to high dimensions, have emerged as promising models for nonlinear time series analysis. Besides, combining physical models helps to further improve the performance of machine learning models.

III. SOLUTION

Based on the preceding discourse, this study has undertaken the selection and investigation of the GBDT and Echo State Network (ESN) models, scrutinizing the alterations in their predictive capabilities when coupled with physical models. The chosen focal point for prediction is the Lorenz64 model, a staple in nonlinear dynamical systems research.

In the "Solution" segment, the exposition commences with an introduction to the Lorenz system, expounding on the methods employed for solving the Lorenz equations. Additionally, it outlines the methodology for establishing an imperfect physical model, thereby imparting fundamental physical insights to the machine learning models. Subsequently, the focus shifts to the evaluation of the GBDT and ESN models, initially utilizing solely observational data for modeling as a benchmark. Subsequent to this, guided by relevant literature, the study delves into two distinct approaches that meld machine learning techniques with the physical model. The first involves the model being trained using both observed data and the predictive outputs of the imperfect model. The second approach

entails the machine learning model learning the discrepancies between the imperfect model and the observed data.

A. Data Acquisition

As mentioned above, this paper chooses the Lorenz system as the research object, so the observation data are obtained by solving the Lorenz equation. The Lorenz system employs fluid convection theory to simulate the movement of a two-dimensional fluid unit, cooled from above and heated from below. The Lorenz63 system comprises three coupled nonlinear equations (Lorenz, 1963):

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = \mathrm{x}(\rho - \mathrm{z}) - \mathrm{y}$$

$$\frac{dz}{dt} = \mathrm{xy} - \beta\mathrm{z}$$

Here, x signifies fluid convection movement, y represents horizontal temperature variation, and z denotes vertical temperature variation. Parameters σ and ρ are associated with the Prandtl and Rayleigh numbers respectively, while parameter β embodies geometric factors. The intricate dynamical behavior and sensitivity to initial conditions make the Lorenz system a challenging prediction target, leading to its wide application in chaotic system prediction studies. Consequently, this model was chosen for analysis in this study.
For obtaining precise system data, the more accurate Runge–Kutta–Fehlberg method was chosen. The method employs a fourth-order higher-order Runge-Kutta scheme and a second-order lower-order Runge-Kutta scheme. An initial value of $(1,1,1)$ was used for solving the equations. The parameter settings were as follows:

$$\sigma = 10.0$$

$$\rho = 28.0$$

$$\beta = \frac{8.0}{3.0}$$

After conducting multiple trial integrations to balance computational cost and precision, a step size of 0.01 was selected. During the initial stage, the system experiences transient behavior, which may introduce instability and hinder model learning and analysis. Hence, the first 5000 data points were discarded after obtaining the actual model, retaining only stable data. The system graph is depicted in Figure 2.

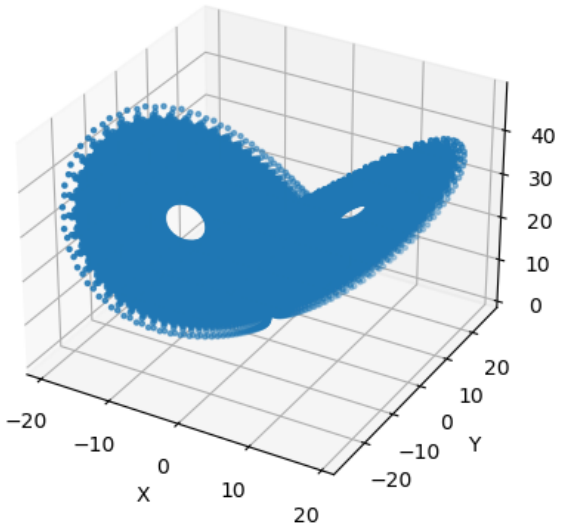

Figure 2. Lorenz Model

To simulate realistic scenarios, obtaining highly frequent observational data is often impractical. Thus, the acquired data was sampled into an observational sequence, with every tenth step being recorded. To emulate inevitable uncertainties in actual measurements, random observational errors were introduced, following a normal distribution with a mean of 0 and a standard deviation of 0.1. The resulting observational sequence is shown in Figure 3.

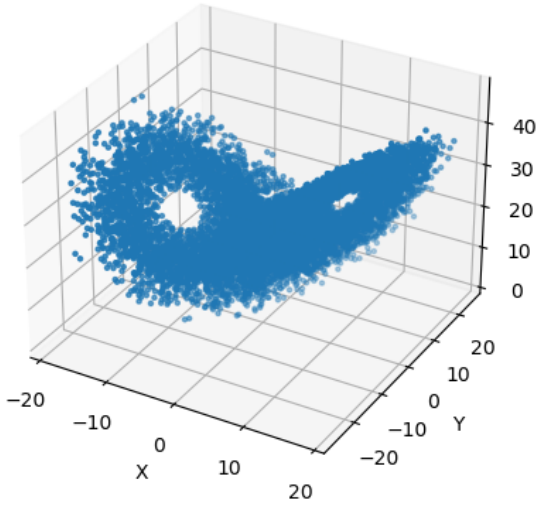

Figure 3. Observed Data

To furnish the model with physical insights, a controlled error was introduced to the Lorenz model, forming an imperfect Lorenz model. This facilitated the observation of the impact of imperfect model errors on learning physical information by the machine learning model. Specifically, the x term was replaced with $c \cdot sin\left(\frac{x}{c}\right)$, c is the error parameter, where, as c approaches positive infinity, $c \cdot sin\left(\frac{x}{c}\right)$ approximates x. This approach engenders controlled errors, resulting in the modified equations:

$$\frac{dx}{dt} = \sigma\left(y - c \cdot sin\left(\frac{x}{c}\right)\right)$$

$$\frac{dy}{dt} = c \cdot sin\left(\frac{x}{c}\right)(\rho - z) - y$$

$$\frac{dz}{dt} = c \cdot sin\left(\frac{x}{c}\right) y - \beta z$$

The RKF method was employed for solving the modified model equations. The process involved using the first observational step as the initial value, followed by predicting each i step using the $i-1$ step observation as the initial value. A step size of 0.01 was employed for ten steps of integration, with the last step being considered as the prediction for step i. This strategy ensured that the system maintained a one-step error. The resulting graph is illustrated in Figure 4, with points deviating from observational data by more than 20 units marked in red.

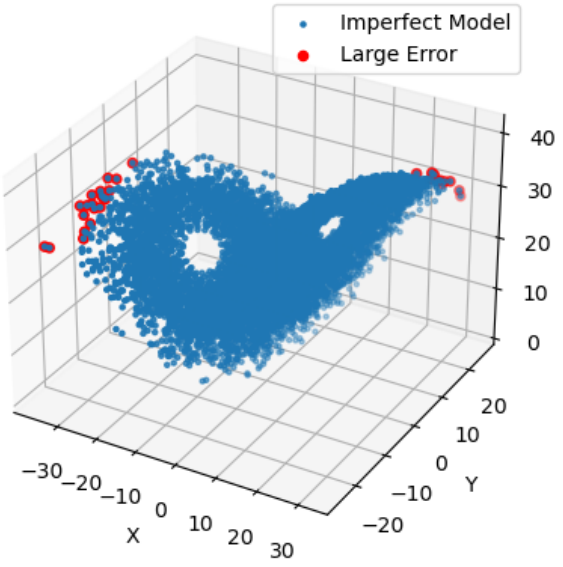

Figure 4. Imperfect Lorenz Model

Observations from Figure 4 reveal that the points with substantial errors (i.e., those significantly diverging from observational data) are concentrated in the outermost region of the x variable of the Lorenz model. Furthermore, these points demonstrate a discernible pattern within this region. This phenomenon suggests a potential correlation between the deviations in the system's outermost x variable and the errors in the imperfect model. In other words, the introduction of the imperfect model appears to provide supplementary information about the behavior of the Lorenz system.

B: GBDT Model

Gradient Boosting Decision Trees (GBDT), proposed by Friedman (2001), is a model that combines statistical and machine learning techniques. Its fundamental concept involves constructing a strong classifier by combining multiple weak classifiers. GBDT employs several shallow-level tree models iteratively, with each tree aiming to learn the residuals of the previous round, thereby achieving efficient prediction.

The procedure begins by determining a constant value γ in the initial stage, such that the loss function $\Sigma_{i=1}^{N} L\left(y_i, F_0\ (x)\right)$ is minimized, where $L(y_i, x_i) = \frac{1}{2}[y_i - F(x_i)]^2$. $F_0\ (x)$ serves as the initial model's prediction output with parameter γ, N is the Training size.

$$F_0\ (x) = \ \text{argmin}_{F_0(x)} \Sigma_{i=1}^{N} L\left(y_i, F_0(x)\right)$$

Where F(x) represents a model composed of all trees, and T(x) represents one of the trees. Subsequently, iterations are performed. For the number of iterations $m = 1: M$ (where M is the number of iterations), calculate the negative gradient direction of residuals y_i^*, which $F(x) = F_{m-1}\ (x)$ represents the model's prediction on sample x_i after the $m-1$times iteration.

$$y_i^* = -\frac{\partial \mathrm{L}(y_i, F(x_i))}{\partial \mathrm{F}(x_i)}\Big|_{F(x)=F_{\mathrm{m}-1}(x)}$$

During each iteration, a new decision tree model $T_{m(x_i)}$ is fitted by minimizing the squared difference between the predicted residuals y_i^* and the output of the new model. β determines the weight of each tree, that is, the learning rate of the model.

$$T_{m(x_i)} = \arg\min_{T_{m(x_i)}} \sum_{i=1}^{N} \left(y_i^* - \beta T_{m(x_i)}\right)^2$$

Within each iteration, a weight factor ω_m is adjusted to minimize the loss function of the sample labels y_i between the current model $\mathrm{F}_{\mathrm{m}-1}(\mathrm{x})$ and the combined output of the newly constructed decision tree model $\mathrm{T}_{\mathrm{m}}(x_i)$. This adjustment aims to update the model weights and reduce the prediction error on the data.

$$\omega_m = \arg\min_{\omega_m} \sum_{i=1}^{N} L\left(y_i, F_{m-1}(x) + \omega_m T_m(x_i; \alpha)\right)$$

Ultimately, by combining the newly constructed decision tree model $T_m(x)$ with the previously iterated model $F_{m-1}(x)$, the updated model $f_m(x)$ is obtained.

$$F_m(x) = F_{m-1}(x) + \omega_m T_m(x)$$

To establish baseline model data, the GBDT model was initially employed to directly predict observational data. The computations were conducted using Kaggle's CPU computing environment, with the model undergoing 500 iterations based on platform capabilities. The optimal model obtained through grid search for hyperparameters yielded a Mean Squared Error (MSE) of 0.0732 for 1024 one-step predictions, as depicted in Figure 5.

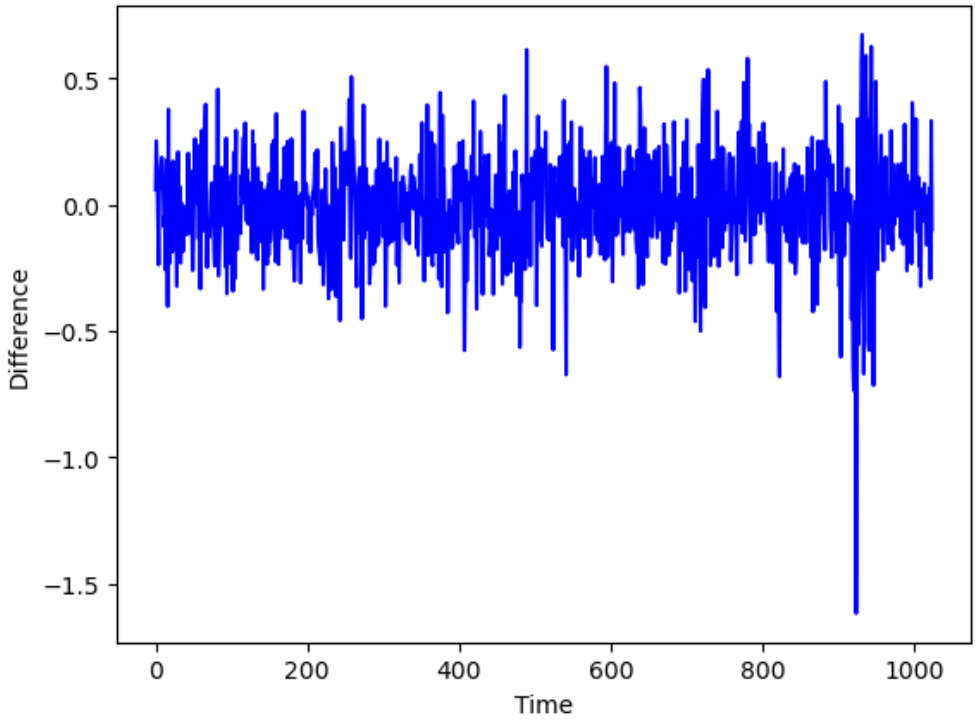

Figure 5. Residual of Observed Data Trained GBDT

Subsequently, two approaches were explored to integrate the physical model. The first method involves predicting the system at step t by inputting both the imperfect model's prediction for step t and the observational data for step $t-1$ to the model. The second approach entails allowing the model to learn and predict the residuals between the imperfect model and observational data. Using $c = 10$ for prediction, the observed data error at step 9500 was 5.352. After grid search for hyperparameters during training, the optimal models using both approaches yielded MSE values of 0.0495 and 0.404 for 1024 one-step predictions. The specific model errors are depicted in Figure 6, where "combined" represents the model that combines predictions from the imperfect model and observational data, while "err" signifies the model trained using error data.

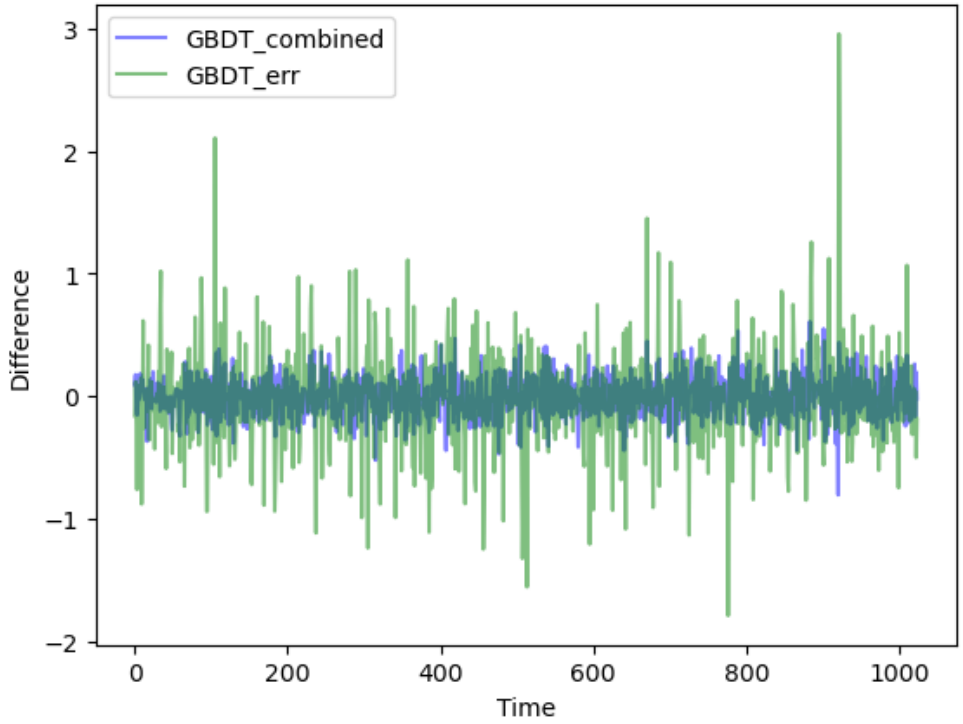

Figure 6. Residual of two Types of Combined Physical Model with GBDT

Results indicated that inputting both the next-step imperfect model prediction and the previous-step observational value effectively enhances predictive capability. However, the approach of learning and predicting the residuals of the imperfect model yielded suboptimal results, even worse than direct observational data prediction. Specific model parameters are detailed in Table 1.

Table 1. Parameter and performance of Best hyperparameter GBDT

GBDT	Estimators	Max Depth	Min Samples Split	Learning Rate	Performance (MSE)
Observation	500	5	5	0.05	0.0732
Combined	500	5	10	0.1	0.0496
err	500	5	2	0.01	0.4038

C. ESN Model

In the preceding sections, we discussed the utilization of the Gradient Boosting Decision Trees (GBDT) model and its integration with physical models. This section delves into the exploration of

another widely acclaimed neural network model that exhibits strong performance in handling chaotic time series – the Echo State Network (ESN).

To maintain the comparability with the previous experiments, we begin by directly training the ESN model solely using observational data. This serves as a baseline for comparison. For consistency between the two models, we set the reservoir size to 1000, based on the computational capabilities of the platform. Although the GBDT algorithm boasts a lower time complexity compared to the ESN, the GBDT's iterative process necessitating waiting for the previous computation results leads to slower processing times. On the contrary, the ESN training process is associated with ridge regression, primarily involving matrix multiplication. Notably, NumPy's matrix multiplication leverages C and Fortran language code at the backend, utilizing functions from the BLAS library. These functions have undergone substantial optimization, employing techniques such as multi-threading and vectorized instructions to expedite computations. Furthermore, NumPy's efficient memory utilization and optimized array data structures further enhance matrix multiplication efficiency. Despite the ESN's relatively higher time complexity, its execution speed is accelerated. Through grid search adjustments of hyperparameters during training, the optimal ESN model achieves a MSE of 0.0268 for 1024 times 1 step error.

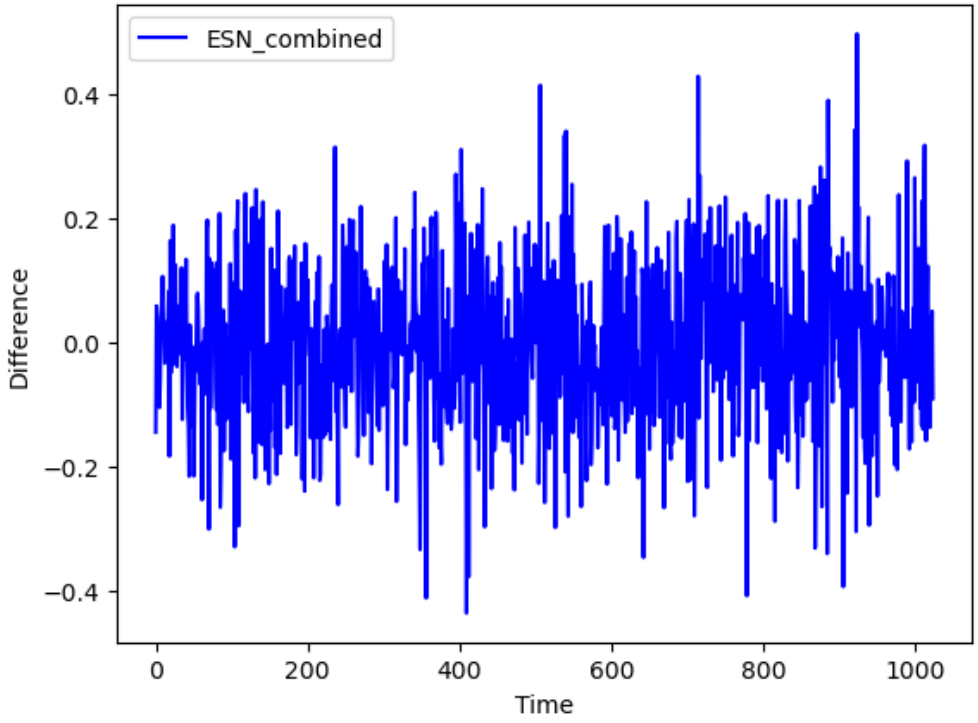

Figure 7. Residual of Observed Data Trained ESN

Subsequently, we extend the research framework from the previous sections to investigate the predictive capabilities of the ESN when combined with a physical model. Retaining the parameter $c = 10$ as a reference point, we further fine-tuned the training process using grid search. The results reveal that the optimal models from both combination methods achieve MSE values of 0.023 and 0.187, respectively. While the first method continues to demonstrate excellent predictive performance, the second method's efficacy remains suboptimal.

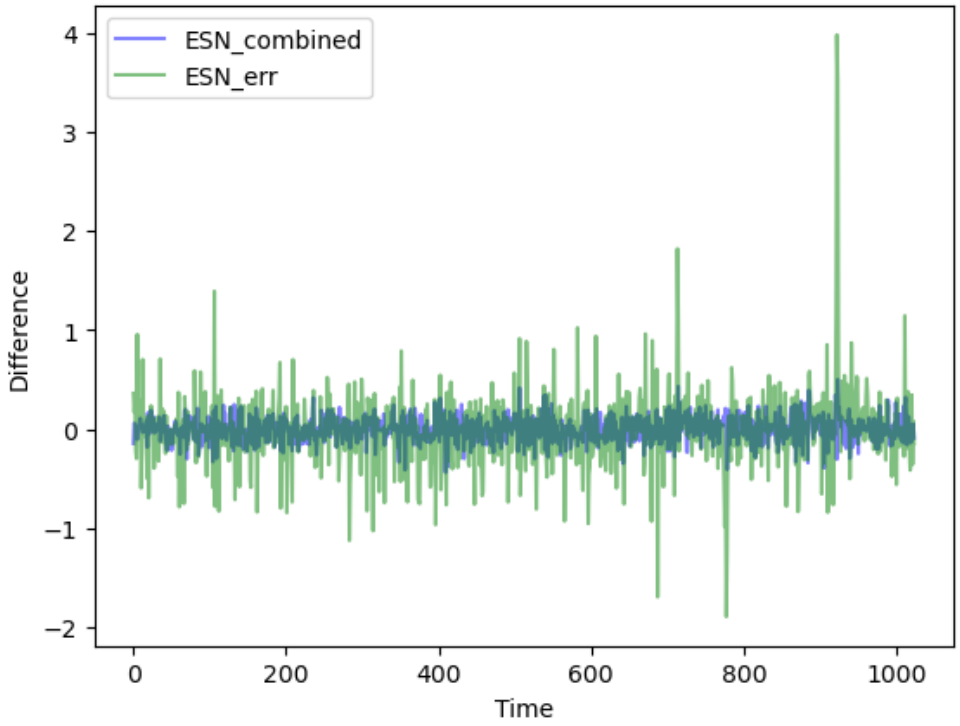

Figure 8. Residual of two Types of Combined Physical Model with ESN

On the whole, a comprehensive assessment suggests that the ESN outperforms GBDT in terms of predictive effectiveness. Detailed information regarding the specific model parameters can be found in Table 2.

Table 2. Parameter and performance of Best hyperparameter ESN

ESN	Reservoir Size	Learning Rate	Spectral Radius	Reg	Performance (MSE)
Observation	1000	1	1	1	0.0268
Combined	1000	0.9	1	1	0.023
err	1000	0.7	1	1	0.1874

D. Robustness Analysis

To further investigate the impact of variations in the performance of imperfect models on the results, this section focuses on four models that integrate physical models. Optimal hyperparameters were selected for each model. The study examines how the 1024 times one-step error changes as parameter c varies from 1 to 14. As c approaches values less than 1, the variation introduced by the imperfect model is modest, resulting in a MSE of around 60 for the 1024 times one-step error. At c=14, the MSE for the 1024 times one-step error reaches 0.867. If c continues to decrease, the imperfect model will converge due to noise.

Consequently, the study further analyzes the influence of varying c from 1 to 14 on the predictive performance of the models, with increments of 1. The results are depicted in Figure 9, with dashed lines representing ESN and GBDT models trained solely on observational data. The intersections of these models with those that combine predictions from imperfect and observational data are highlighted by blue and yellow dots, respectively.

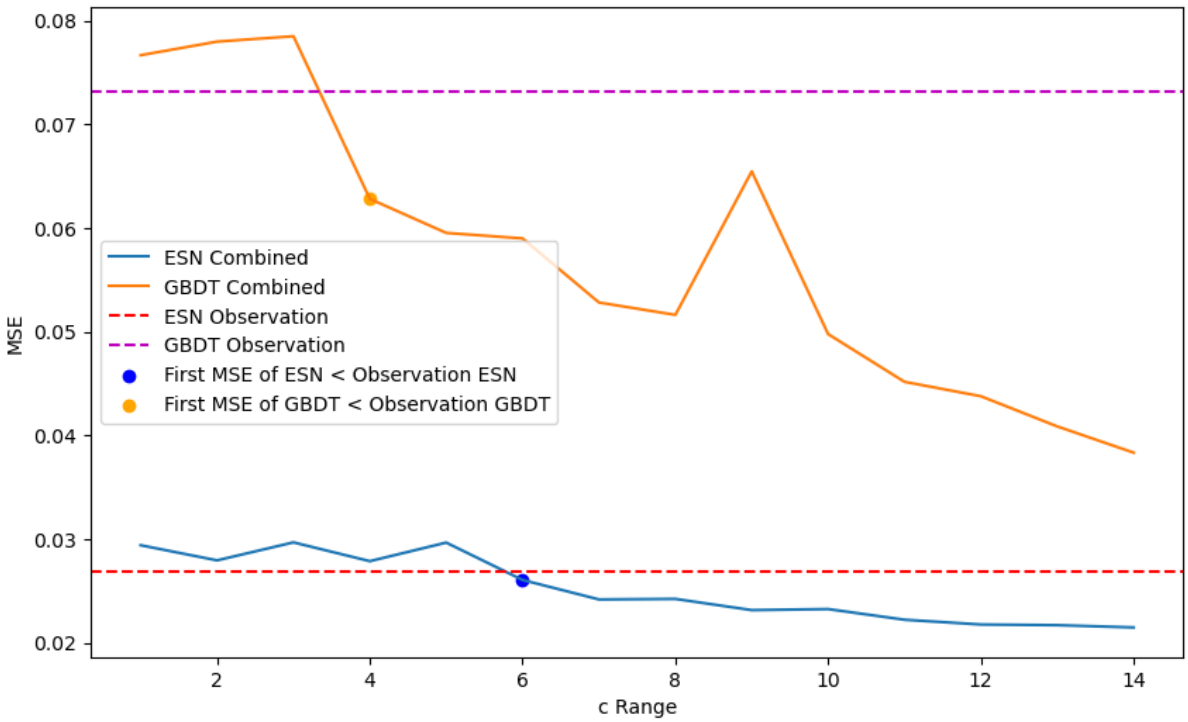

Figure 9. The MSE of Different Models Changes with The Change of c

The findings reveal a consistent trend – throughout the dataset, ESN consistently outperforms the GBDT model. Interestingly, at $c = 4$, the GBDT model effectively learns the physical information of the imperfect model to enhance its performance. In contrast, ESN surpasses the observational data-trained model at $c = 6$. Moreover, with the increase of c, the model effect is continuously optimized, and an outlier value in GBDT may be due to the abnormality of the data.

E. Conclusion

In conclusion, this section has delved into the effectiveness of integrating machine learning models with physical models in the prediction of nonlinear dynamical systems. Using the Lorenz64 model as a benchmark, we employed both GBDT and ESN models to investigate the enhancement of predictive capabilities through two distinct methods of integrating physical models. In the GBDT model experiments, we initially trained the model solely on observational data, obtaining a baseline performance. After 500 iterations, the one-step prediction error (MSE) of the model reached 0.0732. By further combining predictions from physical and observational data in training, our model achieved an MSE reduction to 0.0495 for $c = 10$, representing a 32% improvement. The alternative approach, learning the residuals between the imperfect model and observational data, resulted in an MSE of 0.404, displaying poorer performance compared to direct observational data predictions. In the Echo State Network model experiments, we again trained the model using observational data, yielding a baseline performance with an MSE of 0.0268. Employing the combination of imperfect model predictions and observational data, with a reservoir size of 1000, the model achieved an MSE of 0.023, outperforming the GBDT model and reducing the baseline MSE by 14%. Furthermore, we attempted to learn the residuals between the imperfect model and observational data, resulting in an MSE of 0.187 at $c = 10$, proving less effective than the former approach.

In summary, our research demonstrates that the integration of machine learning models with physical models effectively enhances the accuracy of predicting nonlinear dynamical systems. In both models, ESN outperforms the GBDT model, leveraging the additional information introduced by the imperfect

model to achieve more accurate predictions. Additionally, we conducted sensitivity analysis on the imperfect model parameters, providing valuable insights for further optimizing model performance.

IV. DISCUSSION

In this study, we explored the implementation of two distinct machine learning models, Gradient Boosting Decision Trees (GBDT) and Echo State Networks (ESN), for predicting the chaotic Lorenz system using three different modeling approaches involving observed data and the integration of imperfect physical models. However, the obtained results also raise some questions that require further investigation. Specifically, we need to delve into why the combined use of imperfect model predictions and observed data consistently outperforms training models solely on imperfect model residuals. We also need to explore whether the additional information from the physical model is effectively utilized during the model training process. Furthermore, it is important to analyze the distribution of model residuals and understand how they contribute to the observed outcomes. These questions will be the focus of the Discussion section.

A. Comparison of Two Types of Physical Models

In our experimental analysis, we observed a consistent phenomenon where the performance of models combining imperfect model predictions with observed data was consistently superior to models trained solely on imperfect model residuals. Several factors could potentially explain this phenomenon.

Firstly, a significant contributing factor is the information content. By simultaneously inputting both imperfect model predictions and observed data into the model, we provide a rich source of information from multiple perspectives. This augmentation of information significantly enhances the overall information content available to the model. The combination of imperfect model predictions with historical data not only incorporates the system's history but also infuses it with physical knowledge, resulting in a more informed and comprehensive input. In contrast, when using only residuals, valuable information is discarded, limiting the model's ability to estimate future states accurately.

Secondly, the complexity of the mapping between input and output data also plays a role. The mapping from observed data to the next step is governed by the Lorenz equations, which have a well-defined structure. In contrast, the mapping between successive residuals is inherently complex. Moreover, the residuals are obtained by subtracting two sets of data, which introduces additional intricacies to the mapping. In the case of combining model predictions and observed data, the model benefits from a more straightforward and informative mapping.

Additionally, the direction of information propagation also contributes to this phenomenon. By providing both imperfect model predictions and observed data, the model can receive feedback not only from the past but also from the physical model's predictions of future states. This comprehensive feedback loop contributes to the enhanced predictive performance.

B. Error Analysis

In the result analysis of the machine learning model, a very important link is the residual analysis. Through the analysis of the residual, we can better analyze the fitting of the model and find the abnormality of the model, so as to diagnose or future model improvements and enhancements.

Delving deeper into models combining predictions and observations, we examined the distribution of large error points. We first visualised the baseline ESN model's errors based on observational data. As Figure 10 shows, we highlighted points with errors exceeding 0.7, where larger size indicates greater error.

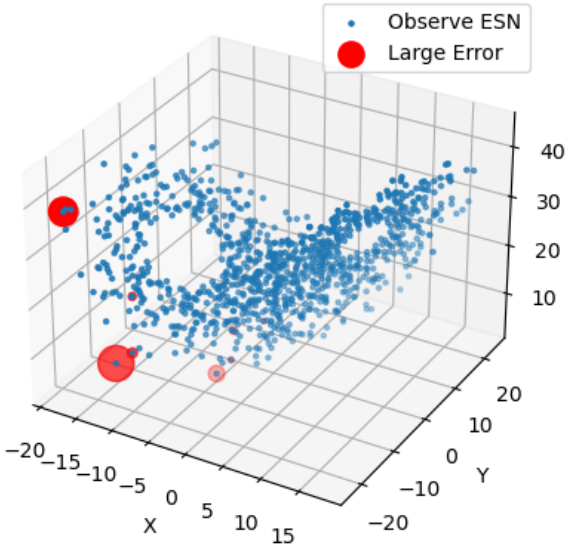

Figure 10. Large Error of Observed Data Trained ESN

Within the Lorenz system's two attractors, we observed a consistent bias of errors towards the left attractor, with significant concentrations of large errors occurring at the outermost trajectory. This raises a pertinent question: why does the model exhibit relatively better learning performance on the right attractor while underperforming on the left? At the outermost extremity of the system trajectory, the y-coordinate experiences the most substantial variation, possibly attributed to the model's diminished capacity to capture rapid y-value changes, thus resulting in larger errors at this outer boundary. However, whether the phenomenon of error concentration around the left attractor is linked to the dynamical characteristics of the attractor or potentially influenced by other factors remains unclear, necessitating further in-depth investigation for elucidation.

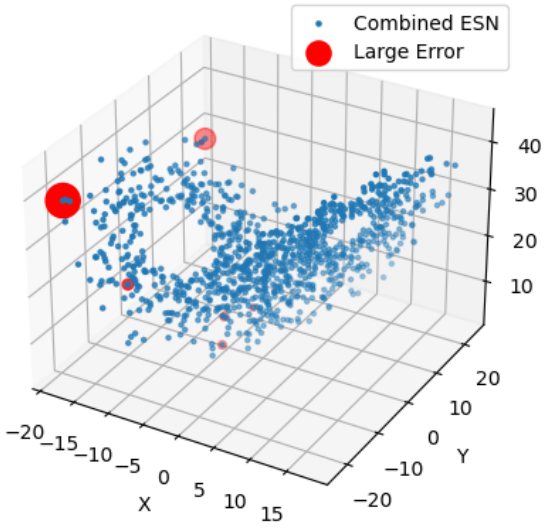

Figure 11. Large Error of Combined Data Trained ESN

Upon introducing the physical model, as depicted in the residual plot (Figure 11), we found a similar trend in the distribution of significant error points. Despite the continued concentration of major errors towards the left attractor, a notable difference emerged from the situation presented in Figure 10. In this case, prominent error points also began to manifest along the positive y-axis. This observation suggests that the incorporation of physical information imparts additional constraints and insights to the model, leading to a more balanced distribution of residuals. However, this stands in contrast to the behavior of imperfect physical models, which, though concentrating major errors at the outermost trajectory, distribute uniformly between the two attractors rather than solely around the left attractor. This suggests that the supplementary physical information does not aid in addressing the issue of inadequate fitting to the left attractor.

Subsequently, we delved into an in-depth examination of the residuals from the benchmark GBDT model. It is noteworthy that the residuals of this model exhibited a uniform distribution across the attractors, extending beyond the outermost region and encompassing the inner regions of the attractors. However, upon closer observation of Figure 5, a predictive failure was evident in the GBDT baseline model, coinciding with the innermost region of the right attractor. Notably, such a significant error within the attractor's interior was not observed in the ESN model. This discrepancy might be attributed to GBDT's inability to capture the temporal sequence in its outputs, training for optimal overall predictions. Given that the outermost region of the trajectory experiences the fastest changes, whereas the innermost region changes at a slower pace, the model's larger errors in handling these distinct cases could arise due to GBDT's lack of temporal sensitivity.

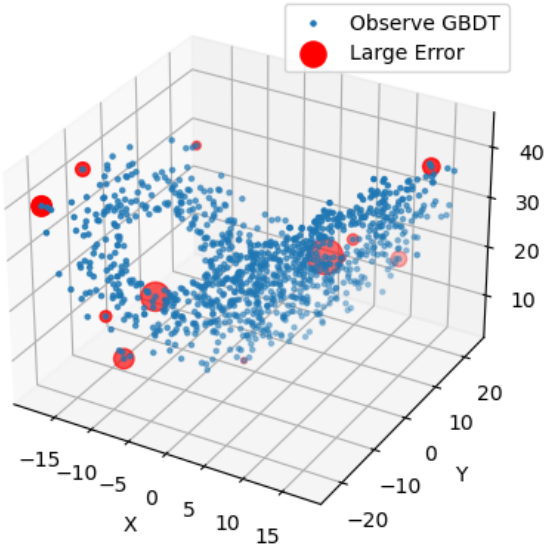

Figure 12. Large Error of Observed Data Trained GBDT

Incorporating physical information resolved this failure. Internal residuals also significantly reduced, leaving just the outermost errors. This further evidence physical information correcting prediction errors and lowering internal residual errors (we highlighted points with errors exceeding 1).

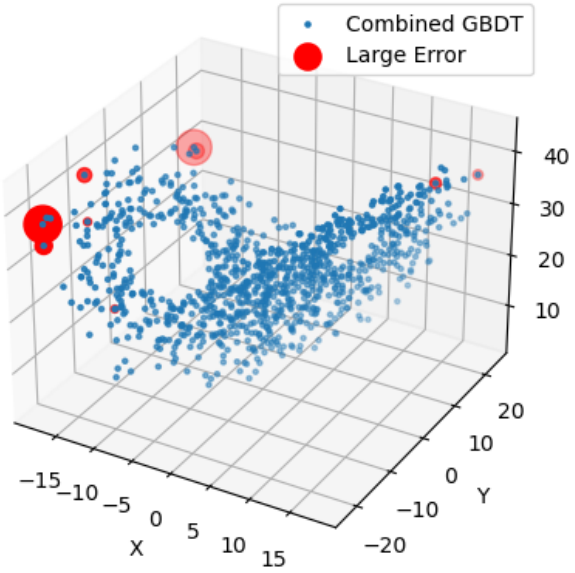

Figure 13. Large Error of Combined Data Trained GBDT

Analysing models combining predictions and observations reveals similarities and differences in how both machine learning models (ESN and GBDT) handle Lorenz system modelling regarding residual distributions. Initially, ESN concentrates residuals on the left attractor while GBDT distributes them uniformly, including exteriors and interiors. However, physical information integration concentrates GBDT residuals only at the outermost areas.

V. CONCLUSION

A. Summary

This study employed the well-known Lorenz system as a case study to investigate the synergistic combination of two distinct machine learning models, Gradient Boosting Decision Trees (GBDT) and Echo State Networks (ESN), with imperfect physical models for predicting complex dynamic systems. To ensure the study's representativeness, we developed a parameter-controlled imperfect Lorenz physical model and generated simulated observational sequences with added noise. This approach balanced the model's physical significance with error control parameters for observing the physical model's impact on machine learning.

Initially, we employed GBDT and ESN models relying solely on observed sequences to predict the Lorenz system, serving as baselines for comparison. The results demonstrated that the ESN model outperformed the GBDT model in terms of predictive accuracy, possibly attributed to ESN's superior handling of dynamic time series data.

Furthermore, we explored two distinct integration methods for combining the physical model with the machine learning models. The first method involved inputting the next-step predictions of the imperfect physical model alongside the previous observed values, while the second method used only the residuals between the imperfect physical model and the observed data. The consistently superior performance of the first method over the second method indicated that the combined input of prediction values and observed data enhanced the predictive accuracy of both GBDT and ESN models compared to solely relying on observed sequences. The improved performance of the former approach is primarily due to the rich information derived from the combination of multiple sources, resembling a form of model fusion, and incorporating prior knowledge from the physical model. Conversely, relying solely on residual information resulted in information loss and less effective performance.

To gain deeper insights into the differences between the two integration methods, we conducted an error analysis. The ESN model's residuals predominantly favored the outer boundary of one attractor, diverging from the anticipated uniform distribution. This intriguing distribution led us to question the underlying causes. Upon introducing physical model residuals (as illustrated in Figure 11), a similar trend persisted, indicating the impact of physical information. Although errors remained biased towards the outer boundary, they were also evident along the positive y-axis, suggesting that the physical model's incorporation imposed additional constraints on the model's error distribution.

Additionally, the analysis of the GBDT model's residuals revealed a more balanced distribution, spanning both attractor boundaries and inner regions. However, the model's predictive performance exhibited a failure case that precisely aligned with the innermost region of an attractor. This failure was rectified when physical model information was integrated, simultaneously mitigating inner attractor errors and correcting the failure.

B. Possible Contribution and Significance

This study has demonstrated from an empirical perspective that the effective combination of machine learning models and physical models can enhance predictive capabilities for complex dynamic systems. On one hand, the physical model provides constraints and insights into system structure and evolution; on the other hand, machine learning models mitigate the limitations of physical models, leading to more accurate state estimation. The synergy between the two approaches surpasses their simple superposition. Importantly, providing only physical model residual information does not yield optimal results, emphasizing the significance of selecting appropriate integration methods.

The study's contribution lies in providing an exemplar of mixed modeling using machine learning and physical models for predicting complex dynamic systems. Through empirical comparison of diverse models and integration strategies, the study lays a foundation for future research in this field and offers insights into selecting suitable hybrid modeling approaches. The combination of machine learning's powerful approximation capabilities and the mechanistic knowledge from physics enables profound understanding and high-precision prediction of nonlinear dynamic system evolution. Further refinement of these models will drive interdisciplinary breakthroughs and the realization of controlling and harnessing chaotic systems.

C. Limitations and Future Directions

Moving forward, there are several areas for potential exploration. Firstly, broader machine learning models, such as recurrent neural networks within deep learning, could be considered. Different model types might exhibit distinct synergistic mechanisms when integrated with physical models. Moreover, constructing imperfect physical models in various forms, simulating different degrees of model biases, can shed light on the impact of model imperfections on the results. Finally, further theoretical analysis of different integration methods, considering information flow and constraint propagation, holds promise for expanding our understanding of the interaction between machine learning and physical models. These avenues await future research to extend our knowledge in this domain.

VI. REFERENCE

Algaba, A., Fernández-Sánchez, F., Merino, M. and Rodríguez-Luis, A.J., 2013. Chen's attractor exists if Lorenz repulsor exists: The Chen system is a special case of the Lorenz system. Chaos: An Interdisciplinary Journal of Nonlinear Science, 23(3).

Balcan, M.F., Blum, A. and Vempala, S., 2006. Kernels as features: On kernels, margins, and low-dimensional mappings. Machine Learning, 65, pp.79-94.

Bengio, Y., Simard, P. and Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks, 5(2), pp.157-166.

Berryman, A.A. and Millstein, J.A., 1989. Are ecological systems chaotic—and if not, why not?. Trends in Ecology & Evolution, 4(1), pp.26-28.
Bradley, E. and Kantz, H., 2015. Nonlinear time-series analysis revisited. Chaos: An Interdisciplinary Journal of Nonlinear Science, 25(9).
Brock, W.A. and Dechert, W.D., 1991. Non-linear dynamical systems: instability and chaos in economics. Handbook of mathematical economics, 4, pp.2209-2235.
Chowdhury, D., Santen, L. and Schadschneider, A., 2000. Statistical physics of vehicular traffic and some related systems. Physics Reports, 329(4-6), pp.199-329.
Connell, J.P., DiMercurio, A., Corbetta, D., Vonk, J. and Shackelford, T., 2017. Dynamic systems theory. Encyclopedia of animal cognition and behavior, pp.1-8.
Elsayed, S., Thyssens, D., Rashed, A., Jomaa, H.S. and Schmidt-Thieme, L., 2021. Do we really need deep learning models for time series forecasting?. arXiv preprint arXiv:2101.02118.
Farchi, A., Bocquet, M., Laloyaux, P., Bonavita, M. and Malartic, Q., 2021. A comparison of combined data assimilation and machine learning methods for offline and online model error correction. Journal of computational science, 55, p.101468.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, pp.1189-1232.
Fu, J., Xiao, D., Fu, R., Li, C., Zhu, C., Arcucci, R. and Navon, I.M., 2023. Physics-data combined machine learning for parametric reduced-order modelling of nonlinear dynamical systems in small-data regimes. Computer Methods in Applied Mechanics and Engineering, 404, p.115771.
Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F. and Liu, Y., 2020. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Computing, 24, pp.16453-16482.
Jaeger, H. and Haas, H., 2004. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. science, 304(5667), pp.78-80.
Jaeger, H. and Haas, H., 2004. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. science, 304(5667), pp.78-80.
Kloeden, P.E. and Mees, A.I., 1985. Chaotic phenomena. Bulletin of Mathematical Biology, 47, pp.697-738.
Llibre, J., Messias, M. and Da Silva, P.R., 2010. Global dynamics of the Lorenz system with invariant algebraic surfaces. International Journal of Bifurcation and Chaos, 20(10), pp.3137-3155.
Lorenz, E.N., 1963. Deterministic nonperiodic flow. Journal of atmospheric sciences, 20(2), pp.130-141.
Lukoševičius, M. and Jaeger, H., 2009. Reservoir computing approaches to recurrent neural network training. Computer science review, 3(3), pp.127-149.
Ren, C., Huang, W. and Gao, D., 2022. Predicting rate of penetration of horizontal drilling by combining physical model with machine learning method in the China Jimusar Oil Field. SPE Journal, pp.1-24.
Rumelhart, D.E., Hinton, G.E. and Williams, R.J., 1986. Learning representations by back-propagating errors. nature, 323(6088), pp.533-536.
Tabor, M. and Weiss, J., 1981. Analytic structure of the Lorenz system. Physical Review A, 24(4), p.2157.
Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V.E., Yagola, A.G.E., Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V. and Yagola, A.G., 1995. Numerical methods for the approximate solution of ill-posed problems on compact sets (pp. 65-79). Springer Netherlands.
Tu, H., Moura, S., Wang, Y. and Fang, H., 2023. Integrating physics-based modeling with machine learning for lithium-ion batteries. Applied Energy, 329, p.120289.
Watson, P.A., 2019. Applying machine learning to improve simulations of a chaotic dynamical system using empirical error correction. Journal of Advances in Modeling Earth Systems, 11(5), pp.1402-1417.
Wen, M., Li, P., Zhang, L. and Chen, Y., 2019. Stock market trend prediction using high-order information of time series. Ieee Access, 7, pp.28299-28308.
Werbos, P.J., 1990. Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 78(10), pp.1550-1560.
Wiggins, S., 2005. The dynamical systems approach to Lagrangian transport in oceanic flows. Annu. Rev. Fluid Mech., 37, pp.295-328.
Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J. and Long, M., 2022. Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186.
Xu, Y., Kohtz, S., Boakye, J., Gardoni, P. and Wang, P., 2022. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. Reliability Engineering & System Safety, p.108900.
Yang, S., Yang, D., Chen, J., Santisirisomboon, J., Lu, W. and Zhao, B., 2020. A physical process and machine learning combined hydrological model for daily streamflow simulations of large watersheds with limited observation data. Journal of Hydrology, 590, p.125206.
Yildiz, I.B., Jaeger, H. and Kiebel, S.J., 2012. Re-visiting the echo state property. Neural networks, 35, pp.1-9.

Zeng, A., Chen, M., Zhang, L. and Xu, Q., 2023, June. Are transformers effective for time series forecasting?. In Proceedings of the AAAI conference on artificial intelligence (Vol. 37, No. 9, pp. 11121-11128).
Zhao, Y., Jiang, C., Vega, M.A., Todd, M.D. and Hu, Z., 2023. Surrogate modeling of nonlinear dynamic systems: a comparative study. Journal of Computing and Information Science in Engineering, 23(1), p.011001.
Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H. and Zhang, W., 2021, May. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 12, pp. 11106-11115).
Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L. and Jin, R., 2022, June. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In International Conference on Machine Learning (pp. 27268-27286). PMLR.