CLIMASIM — Climate Simulation with Scientific Machine Learning

Anonymous submission

Abstract

Climate change manifests as a complex, nonlinear dynamical system characterized by intricate interactions among atmospheric CO_2 concentrations, surface and ocean temperatures, and anthropogenic forcing. While General Circulation Models (GCMs) and their successors, Earth System Models (ESMs), provide comprehensive simulations by incorporating detailed biogeochemical cycles, their computational demands remain prohibitive for rapid climate scenario exploration. Conversely, classical Energy Balance Models (EBMs) offer computational tractability at the expense of predictive accuracy. We present a novel framework that augments EBMs through scientific machine learning, enhancing accuracy while preserving physical interpretability. Specifically, we couple classical energy balance formulations with carbon cycle dynamics and evaluate performance under linearly increasing emission scenarios. Our methodology proceeds systematically: we first investigate Neural Ordinary Differential Equations (Neural ODEs) for climate forecasting, finding limited efficacy. Subsequently, we replace a critical carbon-climate feedback term with a neural network, constructing a Universal Differential Equation (UDE) that achieves error rates below 0.2% across all climate variables for three distinct initializations. To ensure mechanistic transparency, we employ sparse symbolic regression via LASSO, successfully recovering learned dynamics across three initializations and six noise perturbation levels. Comparative benchmarking against statistical baselines (VAR and ARIMA) demonstrates superior forecasting performance in data-scarce regimes with known physical constraints. Our results establish that UDEs enable accurate climate state prediction while symbolic regression maintains interpretability, yielding a computationally efficient framework for rapid climate scenario exploration and mechanistically transparent climate modeling.

Introduction

Climate change represents one of the most pressing challenges of our time, involving complex interactions between atmospheric greenhouse gases, temperature dynamics, and biogeochemical cycles. Understanding and predicting these interactions requires sophisticated modeling approaches that can capture both the physical processes governing Earth's climate system and the nonlinear feedbacks between human activities and environmental responses. General Circulation Models (GCMs), which employ mathematical representations of planetary atmospheric and oceanic circulation based on fundamental physical laws such as the Navier-Stokes equations, have historically served as the foundation for climate modeling (Phillips 1956; Manabe, Smagorinsky, and Strickler 1965; Manabe and Wetherald 1967; Manabe and Bryan 1969). These models incorporate thermodynamic processes including radiation, latent heat exchange, large-scale winds, cloud formation, and ocean-atmosphere interactions. However, the evolution toward Earth System Models (ESMs) has marked a significant advancement by integrating broader Earth system processes beyond purely physical components. ESMs explicitly include chemical and biological processes, comprehensive carbon cycle representations, dynamic vegetation, atmospheric chemistry, and ocean biogeochemistry, creating more complete feedback loops within the climate system (Cox et al. 2000a). While GCMs traditionally rely on predetermined atmospheric compositions as external forcings, ESMs simulate how greenhouse gases and aerosols change endogenously over time in response to both anthropogenic activities and evolving climate conditions (Moon et al. 2025; Song, Scholz, and Lohmann 2025; Schädel et al. 2024; Dittus et al. 2024). Despite their comprehensive nature, both GCMs and ESMs remain computationally intensive, limiting their applicability for rapid scenario exploration and uncertainty quantification(O'Loughlin et al. 2025; Irrgang et al. 2021). In contrast, Energy Balance Models (EBMs), first introduced by Sellers (Sellers 1969), offer simplified representations that sacrifice detail for computational efficiency. EBMs typically focus on energy balance equations relating incoming solar radiation to outgoing terrestrial radiation, often coupled with carbon cycle models to represent biogeochemical feedbacks (Cox et al. 2000b; Stephens et al. 2012; Friedlingstein et al. 2006). While these reduced-complexity models enable efficient exploration of climate scenarios, they may miss important nonlinear dynamics and feedback mechanisms that are crucial for accurate long-term projections.

Related Work

The proposed coupled climate-carbon cycle system builds upon foundational approaches in simplified climate modeling. The DICE model by Nordhaus et al.(Nordhaus 2017) pioneered the integration of two-box energy balance models with coupled carbon cycles, establishing the core feed-

back mechanisms between temperature and carbon dynamics. Goodwin et al.(Goodwin, Williams, and Ridgwell 2007) demonstrated similar coupled frameworks to investigate cumulative carbon emissions and global warming relationships, incorporating temperature-dependent carbon sink feedback analogous to the βTC term in our formulation. Hasselmann et al.(Hasselmann 1997) provided the theoretical foundation for physically-based ordinary differential equations in climate modeling, establishing energy balance models as robust tools for climate change detection and attribution. Joos et al.(Joos et al. 1996) contributed essential insights into carbon cycle representation, validating simplified global carbon cycle formulations that enable efficient coupling with energy balance components.

Scientific Machine Learning (SciML) has emerged as a transformative framework for modeling complex physical systems(Lee and Parish 2021; Vazquez Martinez et al. 2024), integrating physical constraints with neural network expressivity through methodologies including Neural Ordinary Differential Equations (Neural ODEs)(Chen et al. 2018), Universal Differential Equations (UDEs)(Rackauckas et al. 2020), and Sparse Identification of Nonlinear Dynamics (SINDy)(Bournez and Pouly 2020a) for extracting interpretable symbolic expressions.

Recent SciML applications in climate science demonstrate substantial advancement. Choi et al.(Hwang et al. 2021) developed neural diffusion equations combining NODEs with diffusion processes to capture Brownian motion dynamics in climate systems. Bolibar et al.(Bolibar et al. 2023) applied UDEs to glacier ice flow modeling, embedding neural networks within physical equations to discover empirical laws for large-scale glacial processes under changing climate conditions. Neural advection-diffusion equations (NADE)(Choi et al. 2023) further advance physics-informed approaches by incorporating uncertainty modeling within climate frameworks. In addition, neural differential equations have been successfully applied to parameterization of the ocean boundary layer, effectively capturing missing physics in processes driven by surface buoyancy.(Ramadhan 2021)

Motivation

Climate scenario exploration demands computationally efficient frameworks capable of rapid policy evaluation under deep uncertainty. Traditional General Circulation Models and Earth System Models, while comprehensive in capturing atmospheric circulation and biogeochemical feedbacks (Cox et al. 2000a; Moon et al. 2025; Song, Scholz, and Lohmann 2025; Schädel et al. 2024; Dittus et al. 2024), remain computationally prohibitive for iterative scenario analysis. Conversely, Energy Balance Models (Sellers 1969; Cox et al. 2000b; Stephens et al. 2012; Friedlingstein et al. 2006) sacrifice predictive fidelity for computational tractability. Machine learning approaches exhibit excessive data dependence and lack mechanistic transparency—critical deficiencies for regulatory compliance and evidence-based policymaking. Scientific machine learning (Chen et al. 2018; Rackauckas et al. 2020; Bournez and Pouly 2020a; Lee and Parish 2021; Vazquez Martinez et al. 2024) addresses this methodological gap by encoding physical constraints directly into learning architectures, thereby reducing data requirements while preserving interpretability. This study establishes a framework for rapid, interpretable climate scenario exploration by augmenting coupled climate-carbon models with Universal Differential Equations, enabling policymakers to evaluate interventions—carbon taxation, emission caps, adaptation strategies—through mechanistically transparent, computationally efficient simulations that satisfy quantitative risk assessment requirements.

Contributions

This work addresses four fundamental questions in physics-informed climate modeling. First, we establish whether Universal Differential Equations (Rackauckas et al. 2020; Bournez and Pouly 2020a,b) demonstrate superior forecasting accuracy compared to black-box Neural Ordinary Differential Equations (Chen et al. 2018; Dupont, Doucet, and Teh 2019; Massaroli et al. 2020; Yan et al. 2019) through systematic comparison across three random initializations at 1% noise perturbation. Second, we evaluate symbolic recovery performance via Sparse Identification of Nonlinear Dynamics (Bournez and Pouly 2020a) under identical experimental conditions, quantifying the framework's capacity to extract interpretable governing equations. Third, we conduct comprehensive robustness analysis across six noise levels (1-25%), establishing how measurement uncertainty degrades both forecasting accuracy and physics interpretability. Fourth, we benchmark computational efficiency against statistical baselines (VAR and ARIMA), demonstrating resource advantages in data-scarce regimes with known physical constraints. Our framework uniquely achieves dual objectives: accurate climate trajectory prediction and mechanistic discovery of missing physics, enabling researchers to extend incomplete models systematically for rapid scenario exploration.

Methodology

This investigation systematically evaluates three scientific machine learning approaches for climate system identification across a coupled dynamical system governing surface temperature anomaly $T(t)$, deep ocean temperature anomaly $O(t)$, and atmospheric CO_2 concentration $C(t)$. Synthetic datasets are generated via numerical integration with varying observational noise levels to simulate realistic measurement uncertainties. The methodology proceeds hierarchically: Neural Ordinary Differential Equations (Dupont, Doucet, and Teh 2019; Massaroli et al. 2020; Yan et al. 2019) learn complete black-box dynamics from data; Universal Differential Equations (Bournez and Pouly 2020a,b) integrate known partial physics with data-driven components to capture unknown mechanisms; forecasting performance of the UDE framework is benchmarked across various noise regimes against statistical baselines (VAR and ARIMA). Finally, interpretability analysis employs Sparse Identification of Nonlinear Dynamics (Bournez and Pouly 2020a) to recover transparent symbolic expressions of the learned physics. All experiments are implemented in Julia,

with Claude Sonnet (free tier) providing debugging assistance exclusively for syntax verification and package compliance.

Coupled Climate-Carbon System Formulation

We formulate a coupled dynamical system comprising three ordinary differential equations that govern the temporal evolution of the surface temperature anomaly $T(t)$, deep ocean temperature anomaly $O(t)$, and atmospheric CO_2 concentration $C(t)$. This coupled climate-carbon system formulation synthesizes established frameworks from (Nordhaus 2017; Meinshausen, Raper, and Wigley 2011), integrating energy balance dynamics with biogeochemical feedback mechanisms:

$$\frac{dT}{dt} = \frac{1}{C_T}\left(F(C) - \lambda T - \kappa(T - O)\right), \tag{1}$$

$$\frac{dO}{dt} = \frac{1}{C_O}\kappa(T - O), \tag{2}$$

$$\frac{dC}{dt} = E(t) - \beta TC - \gamma C, \tag{3}$$

where the radiative forcing function

$$F(C) = \alpha \log_2\left(\frac{C}{C_0}\right) \tag{4}$$

captures the logarithmic CO_2-forcing relationship, and the remaining parameters govern climate sensitivity (λ), ocean-atmosphere heat exchange (κ), thermal capacities (C_T, C_O), and carbon cycle dynamics (β, γ, $E(t)$) as specified in Table 1.

The surface temperature equation (1) represents energy balance dynamics with radiative forcing as the primary warming driver, counterbalanced by climate feedbacks and oceanic heat transfer mechanisms. The oceanic temperature equation (2) captures deep ocean thermal inertia with characteristically slow response timescales inherent to large thermal reservoirs. The carbon cycle equation (3) combines temperature-independent linear sinks with the temperature-dependent feedback term βTC, quantifying the reduced oceanic CO_2 solubility at elevated temperatures—a critical positive feedback mechanism in the climate system.

The following physical parameters were selected as initial conditions for model integration: climate feedback parameter $\lambda = 1.0$ W m^{-2} K^{-1}, heat exchange coefficient $\kappa = 0.69$ W m^{-2} K^{-1}, surface heat capacity $C_T = 8.0$ W yr m^{-2} K^{-1}, deep ocean heat capacity $C_O = 80.0$ W yr m^{-2} K^{-1}, and temperature-dependent carbon uptake coefficient $\beta = 0.001$ ppm^{-1} K^{-1} yr^{-1}. These values reflect physically plausible magnitudes consistent with simplified climate modeling frameworks. Comprehensive physical interpretation and coupling mechanisms are detailed in Appendix .

Table 1: Physical parameters and initial conditions for climate model simulation

Parameter	Value
Radiative forcing parameter (α)	5.35
Pre-industrial CO_2 concentration (C_0)	280.0 ppm
Linear carbon uptake rate (γ)	0.01 yr^{-1}
Initial surface temperature (T_0)	0.0 K
Initial deep ocean temperature (O_0)	0.0 K
Initial CO_2 concentration ($C(0)$)	280.0 ppm

Data Generation and Preprocessing

Synthetic datasets were generated via Tsit5 Runge-Kutta integration (Tsitouras 2011) applied to Equations (1)–(3), with 15-year training and 50-year validation periods capturing decadal dynamics. Additive Gaussian noise proportional to each variable's intrinsic variability simulated observational uncertainty, where the noise standard deviation for each variable X_i was computed as $\sigma_i = \eta \cdot \text{std}(X_i)$ with noise level η, and anthropogenic emissions prescribed as $E(t) = 8.0 + 0.1t$ Gt C yr^{-1}. Following systematic comparison, Z-score standardization was adopted for optimal convergence. Normalization parameters were computed from clean synthetic data prior to noise injection, preserving statistical fidelity while introducing realistic measurement uncertainty. A safety threshold of 10^{-8} prevented numerical instabilities in near-zero variance cases. To ensure robustness and reproducibility, experiments were conducted across three random initializations with seeds $\mathcal{S} \in \{90, 91, 92\}$, controlling synthetic data generation and neural network parameter initialization. Systematic noise perturbations at levels $\eta \in \{1\%, 5\%, 10\%, 15\%, 20\%, 25\%\}$ of each variable's standard deviation were applied to simulate measurement uncertainty and assess model performance degradation under increasing observational error.

Neural Ordinary Differential Equations

The framework of Neural Ordinary Differential Equations (Neural ODEs) provides a continuous-time representation of dynamical systems through parameterization of the derivative function via neural networks (Dupont, Doucet, and Teh 2019; Massaroli et al. 2020; Yan et al. 2019). This methodology supersedes discrete-time architectural paradigms by establishing a continuous temporal formulation:

$$\frac{d\mathbf{h}}{dt} = f_\theta(\mathbf{h}(t), t), \tag{5}$$

where $\mathbf{h}(t) = [T(t), O(t), C(t)]^T$ constitutes the three-dimensional climate state vector and f_θ represents a fully connected neural network with trainable parameters θ. The adjoint sensitivity method (Ma et al. 2021) facilitates computationally efficient gradient computation through the ODE solver architecture, thereby enabling end-to-end optimization of Neural ODE models for climate time series characterized by extended temporal dependencies and stiff dynamics.

In this work, the neural network f_θ learns the complete dynamical system governing all three climate variables in

a purely data-driven manner, without imposing any mechanistic constraints. This black-box approach serves as a performance benchmark for comparison against physics-informed alternatives. Tables 11, 12, and 13 in the Appendix present the architectural specifications, hyperparameter exploration strategies, and final selected configurations for different random seeds investigated in this study, ensuring reproducibility and robustness assessment across initialization variations. Complete architectural specifications and training configurations across all seeds are provided in Appendix . Experiments were conducted at $\eta = 1\%$ noise level across three random initializations ($\mathcal{S} \in \{90, 91, 92\}$) to evaluate black-box dynamics learning performance under controlled measurement uncertainty.

Universal Differential Equations

The Universal Differential Equation (UDE) methodology establishes a hybrid modeling framework that synergistically integrates mechanistic differential equations with neural network components (Bournez and Pouly 2020a,b). This approach preserves established physical constraints and conservation laws while enabling data-driven identification of unknown or poorly parameterized processes, thereby combining the interpretability of physics-based models with the flexibility of machine learning.

Our UDE implementation strategically substitutes the temperature-dependent carbon uptake term βTC in Equation (3) with a neural network approximation:

$$\frac{dC}{dt} = E(t) - \text{NN}_\phi(T, C) - \gamma C, \tag{6}$$

where $\text{NN}_\phi(T, C)$ constitutes a feedforward neural network with parameters ϕ that learns the complex nonlinear relationship between temperature and CO_2 concentration governing oceanic carbon uptake processes. This formulation maintains the fundamental physical structure of the carbon cycle equation—including known emission inputs and linear sink terms—while providing flexible representational capacity for capturing intricate feedback mechanisms that may be challenging to parameterize from first principles. Tables 20, 21, and 22 in the Appendix present the architectural specifications, hyperparameter exploration strategies, and final selected configurations for different random seeds. For subsequent investigation of symbolic recovery and interpretability across varying noise levels, the configurations detailed in Table 21 were employed. Exhaustive hyperparameter configurations and noise-level analyses are documented in Appendix . Initial validation employed three random seeds ($\mathcal{S} \in \{90, 91, 92\}$) at $\eta = 1\%$ noise level for forecasting accuracy and SINDy symbolic recovery assessment, followed by comprehensive robustness analysis across six noise levels ($\eta \in \{1\%, 5\%, 10\%, 15\%, 20\%, 25\%\}$) using seed $\mathcal{S} = 91$ to quantify performance degradation under increasing observational error.

Statistical Baseline: Autoregressive Integrated Moving Average

In this comparative forecasting analysis, we employ univariate ARIMA models(Box and Jenkins 1976) as a classical statistical baseline for climate variable prediction. ARIMA(p, d, q) models capture temporal dependencies through autoregressive terms of order p, differencing operations of order d for achieving stationarity, and moving average components of order q for modeling residual correlations. We implement three independent ARIMA models, one for each climate variable: surface temperature anomaly, deep ocean temperature anomaly, and atmospheric CO_2 concentration.

Hyperparameter optimization is performed via exhaustive grid search over the parameter space detailed in Table 43, evaluating all combinations using the corrected Akaike Information Criterion (AICc) as the selection metric. The AICc criterion penalizes model complexity while rewarding goodness-of-fit, making it particularly suitable for time series with limited sample sizes. Each model is fitted independently on the training period using maximum likelihood estimation with mean centering enabled to account for non-zero means in the climate anomaly data. The optimal configuration for each variable is determined by minimizing AICc across 48 candidate model specifications, after which multi-step-ahead forecasts are generated for the validation period via recursive prediction, where forecasted values serve as inputs for subsequent time steps. Grid search results and forecasting metrics are comprehensively presented in Appendix . All ARIMA experiments utilized random seed $\mathcal{S} = 91$ with observational noise level $\eta = 1\%$ to enable direct performance comparison against alternative methodologies under controlled stochastic conditions.

Statistical Baseline: Vector Autoregression

For multivariate time series forecasting, we employ a Vector Autoregression (VAR) model(Sims 1980; Litterman 1986; Basu and Michailidis 2015) that simultaneously captures the dynamic interdependencies among surface temperature, deep ocean temperature, and CO_2 concentration. Unlike univariate approaches that model each variable in isolation, the VAR(p) framework expresses each variable as a linear function of its own lagged values and the lagged values of all other variables in the system, thereby explicitly exploiting cross-variable temporal correlations and Granger causality structures inherent in coupled climate dynamics.

The optimal lag order p is determined through systematic evaluation of information criteria across candidate models, as specified in Table 46. We compute the Akaike Information Criterion (AIC)(Akaike 1973), Bayesian Information Criterion (BIC)(Schwarz 1978), Hannan-Quinn Criterion (HQC)(Hannan and Quinn 1979), and Final Prediction Error (FPE) for each lag configuration. The model that minimizes BIC is ultimately selected due to its stronger penalty for model complexity, promoting parsimony—a desirable property when working with relatively short time series. The VAR system is estimated via ordinary least squares (OLS) regression with a constant term included to accommodate non-zero means, and model stability is rigorously verified through eigenvalue analysis of the companion matrix to ensure all roots lie within the unit circle. Multi-horizon forecasts are generated recursively using the estimated coefficient matrices, propagating uncertainty through the predic-

tion horizon. Mathematical formulation and lag selection procedures are rigorously detailed in Appendix . The VAR framework was evaluated under random seed $\mathcal{S} = 91$ with $\eta = 1\%$ Gaussian noise perturbations to establish baseline statistical forecasting performance under identical experimental conditions.

Symbolic Regression and Interpretability Analysis

The SINDy (Sparse Identification of Nonlinear Dynamics) algorithm enables the extraction of human-interpretable symbolic expressions from opaque neural network representations (Bournez and Pouly 2020a), thereby transforming black-box predictions into transparent mathematical formulations. Given normalized state variables (T_{norm}, C_{norm}) and neural network outputs $\mathbf{y}_{NN}$ representing the learned carbon uptake mechanism, the algorithm constructs a polynomial basis library:

$$\Phi(T_{norm}, C_{norm}) = [TC, T^2C, TC^2, T^2C^2] \quad (7)$$

This interpretability-focused regression problem is formulated using non-negative least squares optimization, solved via the NonNegLeastSquares.jl package:

$$\min_{\boldsymbol{\xi}} \|\Phi\boldsymbol{\xi} - \mathbf{y}_{NN}\|_2^2 \quad \text{subject to} \quad \boldsymbol{\xi} \geq 0 \quad (8)$$

A coefficient thresholding procedure (threshold = 0.001) identifies dominant terms while eliminating spurious contributions, yielding a parsimonious and interpretable symbolic representation of the complex temperature-carbon feedback dynamics originally embedded within the neural network architecture. Non-negative least squares algorithm implementation is explicated in Appendix .

Results

Forecasting Performance

The sequential Adam-AdamW optimization framework achieved exceptional loss reductions exceeding 99.57% for UDE configurations (Tables 23–25) and surpassing 99.997% for Neural ODE architectures (Tables 14–16) under 1% noise conditions, as illustrated in Figures 11–17. Systematic noise augmentation experiments (Tables 26–30) demonstrated progressive degradation from 99.56% to 91.05% loss reduction as noise levels increased from 5% to 25%, confirming the optimizer combination's robustness. Adam exhibited more stable convergence behavior, whereas AdamW showed increased volatility and instability in later training stages. L2 error trajectories (Figures 13–18, Appendix) exhibited monotonic growth throughout the 50-year forecast horizon for Neural ODE architectures, with cumulative error amplification indicating progressive deviation from ground-truth dynamics across all seed initializations.

Table 2: Forecasting Performance: Neural ODE vs. UDE

Variable	Model	Mean % Error
Surface Temperature	Neural ODE	12.45
	UDE	0.14
Ocean Temperature	Neural ODE	64.08
	UDE	0.11
CO_2 Concentration	Neural ODE	5.17
	UDE	0.11

Table 3: Training Efficiency: Neural ODE vs. UDE

Metric	Neural ODE	UDE
Mean Iterations	40,002	17,002
Mean Loss Reduction (%)	99.999	99.846

UDE demonstrates superior forecasting performance compared to Neural ODE across all variables (Table 2), achieving substantially lower mean percentage errors: surface temperature (0.14% vs. 12.45%), ocean temperature (0.11% vs. 64.08%), and CO_2 concentration (0.11% vs. 5.17%). Furthermore, UDE exhibits enhanced computational efficiency, requiring 57.5% fewer iterations while maintaining comparable loss reduction (Table 3).

Table 4: Forecasting Errors at $t = 50$ Years: Neural ODE, UDE, and VAR

Climate Variable	Neural ODE	UDE	VAR
Surface Temp. (°C)	0.40	0.31	22.41
Ocean Temp. (°C)	25.83	0.20	19.60
CO_2 Conc. (ppm)	9.88	0.23	6.51

Table 5: Forecasting Errors at $t = 50$ Years: Neural ODE, UDE, and ARIMA

Climate Variable	Neural ODE	UDE	ARIMA
Surface Temp. (°C)	0.40	0.31	99.25
Ocean Temp. (°C)	25.83	0.20	70.34
CO_2 Conc. (ppm)	9.88	0.23	1.62

Table 6: Computational Resources: Neural ODE, UDE, and ARIMA

Resource Metric	Neural ODE	UDE	ARIMA
Parameters	12931	8577	10
Training Iterations	40000	11000	144
Training Time (s)	1036.99	193.35	7.52
Time/Iter. (ms)	25.92	17.58	52.25

Table 7: Computational Resources: Neural ODE, UDE, and VAR

Resource Metric	Neural ODE	UDE	VAR
Parameters	12931	8577	21
Training Iterations	40000	11000	2
Training Time (s)	907.33	208.81	0.99
Time/Iter. (ms)	22.68	18.98	495.04

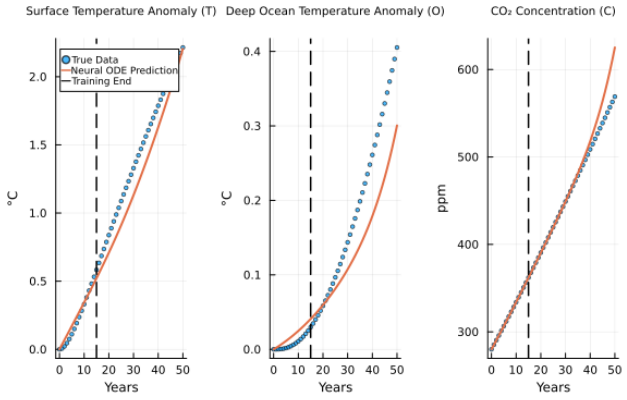

Figure 1: Neural ODE Climate Variable Predictions

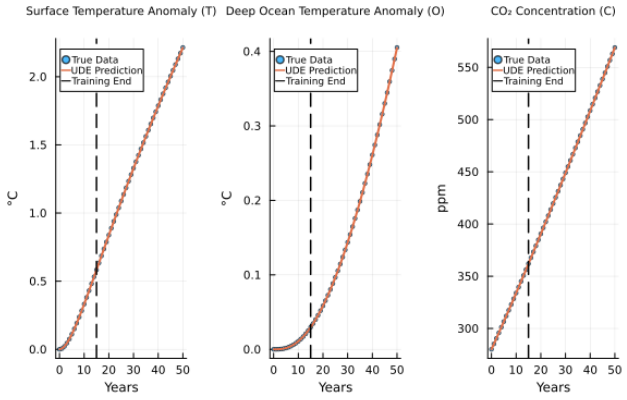

Figure 2: Universal Differential Equation Climate Variable Predictions

Table 8: UDE Forecasting Errors Across Noise Levels

Noise (%)	T Error (%)	O Error (%)	C Error (%)
1	0.04	0.08	0.007
5	0.29	0.33	0.17
10	0.52	0.59	0.33
15	0.25	0.56	0.08
20	0.42	0.74	0.21
25	50.27	33.23	32.05

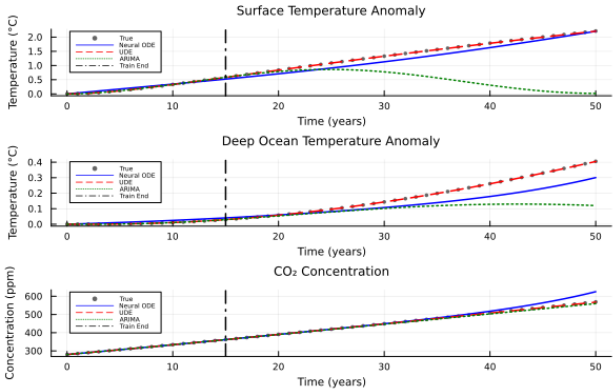

Figure 3: Comparative Forecasting Performance: ARIMA, Neural ODE, and UDE

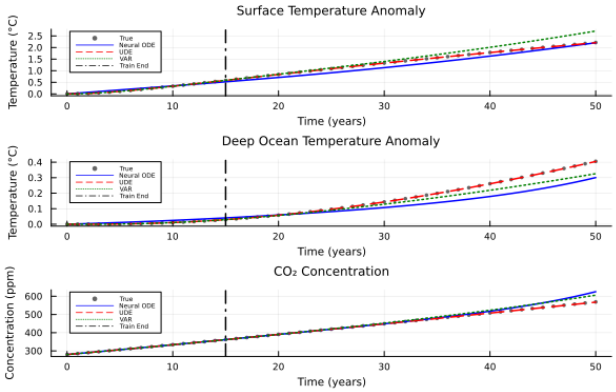

Figure 4: Comparative Forecasting Performance: VAR, Neural ODE, and UDE

The Universal Differential Equation (UDE) architecture demonstrates superior predictive performance across all climate variables (Table 4, Table 5), achieving percentage errors below 0.31% for surface temperature and 0.23% for CO_2 concentration at $t = 50$ years. Comparative analysis against statistical baselines reveals that Neural ODE, ARIMA, and VAR models exhibit substantial gaps between true and predicted values across all three climate variables (Figures 3–4), whereas UDE predictions maintain precise alignment with ground truth dynamics, demonstrating no observable discrepancy between true and predicted trajectories. While statistical baselines (VAR, ARIMA) exhibit computational efficiency with minimal parameters (Table 6, Table 7), they incur substantial prediction errors exceeding 19% for temperature variables. The UDE framework achieves optimal accuracy-efficiency tradeoff, requiring 33.6% fewer parameters and 72.5% less training time than Neural ODE while maintaining significantly lower percentage errors across all prognostic variables.

As demonstrated in Table 8, UDE predictions maintain remarkably low percentage errors across noise levels (0.25–0.74% for 5–20% noise), with catastrophic degradation only at 25% noise (32–50% error).

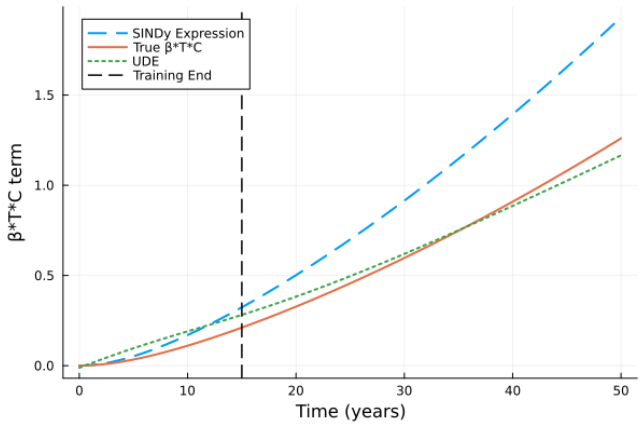

Figure 6: Symbolic Recovery of Temperature-Carbon Feedback Term at 15% Noise

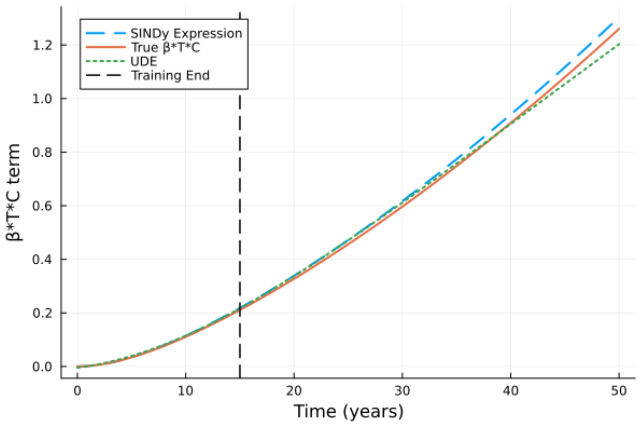

Figure 7: Symbolic Recovery of Temperature-Carbon Feedback Term at 1% Noise

Interpretability Analysis

The symbolic recovery analysis (Figures 7–9) reveals progressive deviation of SINDy predictions from both ground truth and UDE forecasts as noise intensifies. SINDy successfully recovered the correct functional form $\beta \cdot T \cdot C$ across all noise regimes up to 20%, demonstrating robust symbolic identification capabilities.

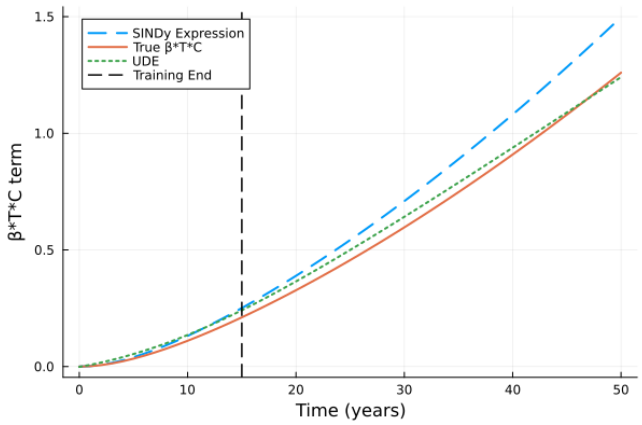

Figure 5: Symbolic Recovery of Temperature-Carbon Feedback Term at 5% Noise

Table 9: Symbolic Regression Performance at 1% Noise

Metric	Value	Unit
Average Relative Error	25.22	%
Average Coefficient (β)	0.00077	–

However, at 25% noise (Figure 10, Table 8), interpretability undergoes complete breakdown, with SINDy exhibiting precipitous R^2 collapse to 0.4028 and 100% relative error in parameter estimation (Table 10), effectively rendering symbolic identification infeasible beyond this critical threshold.

While ARIMA yielded closed-form equations with inadequate forecasting fidelity and Neural ODEs remained entirely black-box architectures precluding mechanistic interpretation, only the UDE framework achieved simultaneous predictive accuracy and transparent symbolic recovery of governing dynamics through SINDy-based interpretability analysis.

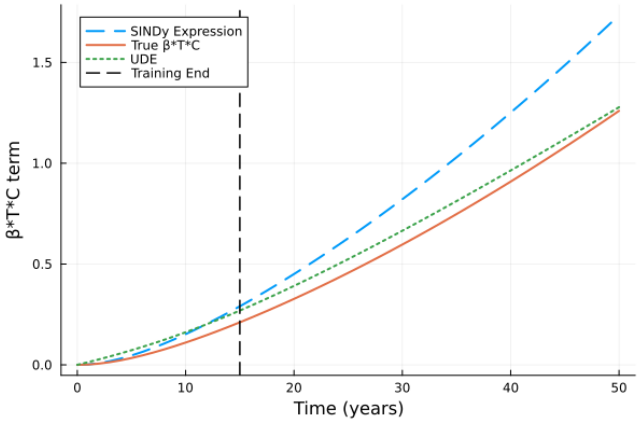

Figure 8: Symbolic Recovery of Temperature-Carbon Feedback Term at 10% Noise

Table 10 quantifies SINDy performance metrics relative to the ground truth parameter value $\beta = 0.001$, demonstrating systematic degradation in both goodness-of-fit (R^2) and parametric accuracy as observational noise escalates from 1% to 25%. Across three random seeds, SINDy successfully recovered the correct functional form ($T \times C$) in all cases with a 100% recovery rate (Table 9), achieving an average relative error of 25.22% at 1% noise level.

Conclusion

Neural Ordinary Differential Equations demonstrated inferior forecasting performance relative to Universal Dif-

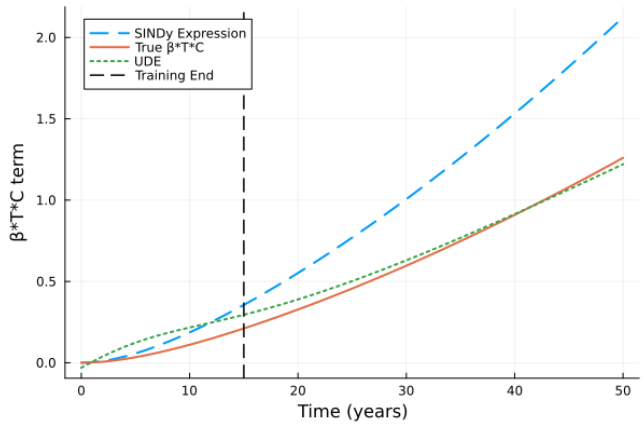

Figure 9: Symbolic Recovery of Temperature-Carbon Feedback Term at 20% Noise

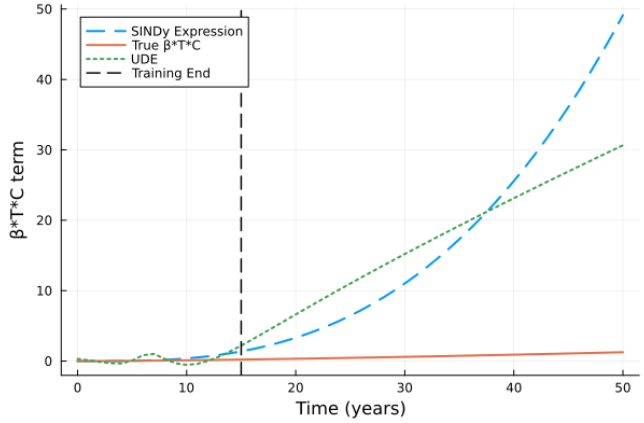

Figure 10: Symbolic Recovery of Temperature-Carbon Feedback Term at 25% Noise

Table 10: Symbolic Regression Performance Across Noise Levels

Noise(%)	SINDy R^2	β Relative Error (%)
1	0.9985	3.45
5	0.9868	18.86
10	0.9558	37.64
15	0.8820	53.26
20	0.7812	68.58
25	0.4028	100.00

ferential Equations across all climate variables, exhibiting substantially elevated mean percentage errors and progressive error amplification throughout the extended forecast horizon. Despite achieving comparable loss reduction during training, NODEs required markedly greater computational resources, necessitating over twice the iterations and training time compared to UDEs while maintaining significantly larger parametrization. This performance disparity underscores NODEs' fundamental limitation in capturing complex climate dynamics without explicit incorporation of physical constraints.

Conversely, UDEs successfully achieved high-fidelity approximations of climate system evolution, maintaining remarkably low percentage errors across all prognostic variables under moderate noise conditions. The architecture demonstrated exceptional robustness across escalating noise regimes, preserving predictive accuracy until catastrophic degradation at critical noise thresholds. This resilience reflects UDEs' privileged access to complete differential equation structures and comprehensive input data, enabling superior learning capacity compared to purely data-driven approaches.

Symbolic regression analysis revealed progressive interpretability degradation as observational noise intensified, with SINDy maintaining functional form recovery across moderate noise levels despite deteriorating goodness-of-fit and parametric accuracy. The interpretability breakdown at extreme noise conditions suggests fundamental limitations of symbolic identification when operating solely on neural network outputs rather than complete system dynamics. Nevertheless, successful term recovery under realistic noise scenarios demonstrates substantial promise for transforming opaque learned components into explicit mathematical formulations.

Comparative evaluation against statistical baseline methodologies confirmed that while VAR and ARIMA models exhibit superior computational parsimony with minimal parametrization and negligible training requirements, they incur substantial prediction errors across all climate variables, rendering them unsuitable for high-fidelity forecasting applications. This performance disparity validates the fundamental advantage of incorporating physical law constraints within machine learning frameworks. The UDE architecture's optimal accuracy-efficiency tradeoff positions it as particularly advantageous for rapid climate scenario exploration and data-constrained forecasting applications.

Limitations and Future Directions

This study's primary limitation involves synthetic data usage, restricting immediate real-world applicability. For future research, we plan to validate these approaches using authentic observational data and expand the model framework to incorporate additional variables such as stratospheric temperature, oceanic pH, methane concentration, ice sheet volume, and vegetation biomass with appropriate coupling terms and governing equations.

References

Akaike, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In Petrov, B. N.; and Csáki, F., eds., *Second International Symposium on Information Theory*, 267–281. Budapest: Akadémiai Kiadó.

Basu, S.; and Michailidis, G. 2015. Regularized estimation of high-dimensional vector autoregressive models. *Journal of Multivariate Analysis*, 137: 20–34.

Bolibar, J.; Sapienza, F.; Maussion, F.; Lguensat, R.; Wouters, B.; and Pérez, F. 2023. Universal differential equa-

tions for glacier ice flow modelling. *Geoscientific Model Development*, 16(22): 6671–6687.

Bournez, O.; and Pouly, A. 2020a. A universal ordinary differential equation. *Logical Methods in Computer Science*, 16.

Bournez, O.; and Pouly, A. 2020b. A universal ordinary differential equation. *Logical Methods in Computer Science*, 16.

Box, G. E. P.; and Jenkins, G. M. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 2 edition.

Chen, R. T.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural ordinary differential equations. *Advances in neural information processing systems*, 31.

Choi, H.; Choi, J.; Hwang, J.; Kim, Y.-W.; Lee, S.; Kim, S. H.; and Ko, K.-E. 2023. Climate modeling with neural advection–diffusion equation. *Knowledge and Information Systems*, 65(6): 2403–2427.

Cox, P. M.; Betts, R. A.; Jones, C. D.; Spall, S. A.; and Totterdell, I. J. 2000a. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408(6809): 184–187.

Cox, P. M.; Betts, R. A.; Jones, C. D.; Spall, S. A.; and Totterdell, I. J. 2000b. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408(6809): 184–187.

Dittus, A. J.; Hawkins, E.; Wilcox, L. J.; and et al. 2024. Reversal of Projected European Summer Precipitation Decline in a Stabilizing Climate. *Geophysical Research Letters*, 51(e2023GL107448).

Dupont, E.; Doucet, A.; and Teh, Y. W. 2019. Augmented neural odes. *Advances in neural information processing systems*, 32.

Friedlingstein, P.; Cox, P. M.; Betts, R. A.; Bopp, L.; von Bloh, W.; Brovkin, V.; Cadule, P.; Doney, S. C.; Eby, M.; Fung, I.; Bala, G.; John, J.; Jones, C. D.; Joos, F.; Kato, T.; Kawamiya, M.; Knorr, W.; Lindsay, K.; Matthews, H. D.; Raddatz, T.; Rayner, P. J.; Reick, C.; Roeckner, E.; Schnitzler, K.-G.; Schnur, R.; Strassmann, K.; Weaver, A. J.; Yoshikawa, C.; and Zeng, N. 2006. Climate-Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison. *Journal of Climate*, 19(14): 3337–3353.

Goodwin, P.; Williams, R. G.; and Ridgwell, A. 2007. Sensitivity of climate to cumulative carbon emissions. *Geophysical Research Letters*, 34(18).

Hannan, E. J.; and Quinn, B. G. 1979. The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society, Series B (Methodological)*, 41(2): 190–195.

Hasselmann, K. 1997. Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics*, 13(9): 601–611.

Hwang, J.; Choi, J.; Choi, H.; Lee, K.; Lee, D.; and Park, N. 2021. Climate Modeling with Neural Diffusion Equations. In *Proceedings - 21st IEEE International Conference on Data Mining, ICDM 2021*, 230–239. Institute of Electrical and Electronics Engineers Inc.

Irrgang, C.; Boers, N.; Sonnewald, M.; Barnes, E. A.; Kadow, C.; Staneva, J.; Pohlmann, H.; Agarwal, N.; Brune, S.; Müller, A.; and Baehr, J. 2021. Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. *Nature Machine Intelligence*, 3: 667–674.

Joos, F.; Bruno, M.; Siegenthaler, U.; Stocker, T. F.; Quéré, C. L.; and Plattner, G.-K. 1996. An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake. *Tellus B: Chemical and Physical Meteorology*, 48(3): 397–417.

Lee, K.; and Parish, E. J. 2021. Parameterized neural ordinary differential equations: Applications to computational physics problems. *Proceedings of the Royal Society A*, 477(2253): 20210162.

Litterman, R. B. 1986. Forecasting with Bayesian Vector Autoregressions–Five Years of Experience. *Journal of Business & Economic Statistics*, 4(1): 25–38.

Ma, Y.; Dixit, V.; Innes, M. J.; Guo, X.; and Rackauckas, C. 2021. A Comparison of Automatic Differentiation and Continuous Sensitivity Analysis for Derivatives of Differential Equation Solutions. In *2021 IEEE High Performance Extreme Computing Conference (HPEC)*, 1–9. IEEE.

Manabe, S.; and Bryan, K. 1969. Climate and the Ocean Circulation. *Monthly Weather Review*, 97(11): 806–827.

Manabe, S.; Smagorinsky, J.; and Strickler, R. F. 1965. Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly Weather Review*, 93(12): 769–798.

Manabe, S.; and Wetherald, R. T. 1967. Thermal Equilibrium of the Atmosphere with a Given Distribution of Relative Humidity. *Journal of the Atmospheric Sciences*, 24(3): 241–259.

Massaroli, S.; Poli, M.; Park, J.; Yamashita, A.; and Asama, H. 2020. Dissecting neural odes. *Advances in Neural Information Processing Systems*, 33: 3952–3963.

Meinshausen, M.; Raper, S. C. B.; and Wigley, T. M. L. 2011. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration. *Atmospheric Chemistry and Physics*, 11(4): 1417–1456.

Moon, J.-Y.; Streffing, J.; Lee, S.-S.; Semmler, T.; Andrés-Martínez, M.; Chen, J.; Cho, E.-B.; Chu, J.-E.; Franzke, C. L. E.; Gärtner, J. P.; Ghosh, R.; Hegewald, J.; Hong, S.; Kim, D.-W.; Koldunov, N.; Lee, J.-Y.; Lin, Z.; Liu, C.; Loza, S. N.; Park, W.; Roh, W.; Sein, D. V.; Sharma, S.; Sidorenko, D.; Son, J.-H.; Stuecker, M. F.; Wang, Q.; Yi, G.; Zapponini, M.; Jung, T.; and Timmermann, A. 2025. Earth's future climate and its variability simulated at 9 km global resolution. *Earth System Dynamics*, 16: 1103–1134.

Nordhaus, W. D. 2017. Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 114(7): 1518–1523.

O'Loughlin, R. J.; Lguensat, R.; Labe, D.; and O'Brien, T. A. 2025. Moving beyond post hoc explainable artificial intelligence: a perspective paper on lessons learned from dynamical climate modeling. *Geoscientific Model Development*, 18: 787–828.

Phillips, N. A. 1956. The general circulation of the atmosphere: A numerical experiment. *Quarterly Journal of the Royal Meteorological Society*, 82(352): 123–164.

Rackauckas, C.; Ma, Y.; Martensen, J.; Warner, C.; Zubov, K.; Supekar, R.; Skinner, D.; Ramadhan, A.; and Edelman, A. 2020. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*.

Ramadhan, A. 2021. *Data-driven ocean modeling using neural differential equations*. Ph.D. thesis, Massachusetts Institute of Technology.

Schwarz, G. E. 1978. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2): 461–464.

Schädel, C.; Rogers, B.; Lawrence, D.; and et al. 2024. Earth system models must include permafrost carbon processes. *Nature Climate Change*, 14: 114–116.

Sellers, W. D. 1969. A Global Climatic Model Based on the Energy Balance of the Earth-Atmosphere System. *Journal of Applied Meteorology*, 8: 392–400.

Sims, C. A. 1980. Macroeconomics and Reality. *Econometrica*, 48(1): 1–48.

Song, P.; Scholz, P.; and Lohmann, G. 2025. Regional conditions determine thresholds of accelerated Antarctic basal melt in climate projection. *Nature Climate Change*, 15: 1–7.

Stephens, G. L.; Li, J.; Wild, M.; van den Broeke, M. R.; Cazenave, A.; Edwards, N. R.; Gettelman, A.; Ghan, S.; Kaser, G.; Kato, S.; et al. 2012. An update on Earth's energy balance in light of the latest global observations. *Nature Geoscience*, 5(10): 691–696.

Tsitouras, C. 2011. Runge–Kutta pairs of order 5 (4) satisfying only the first column simplifying assumption. *Computers & Mathematics with Applications*, 62(2): 770–775.

Vazquez Martinez, R.; Abhijit Dandekar, R.; Dandekar, R.; and Panat, S. 2024. A comparative study of NeuralODE and Universal ODE approaches to solving Chandrasekhar White Dwarf equation. *arXiv e-prints*, arXiv–2410.

Yan, H.; Du, J.; Tan, V. Y.; and Feng, J. 2019. On robustness of neural ordinary differential equations. *arXiv preprint arXiv:1910.05513*.

Appendix

Model formulation

This subsection provides comprehensive physical interpretation of the coupled energy balance and carbon cycle dynamics, including radiative forcing mechanisms, thermal reservoir interactions, ocean-atmosphere heat exchange, and temperature-dependent biogeochemical feedback pathways governing the three-state climate system.

Our proposed model represents a simplified climate system based on energy balance principles, employing a box model approach that discretizes the climate into interconnected reservoirs without the computational complexity of full-scale General Circulation Models. The model tracks three fundamental state variables as deviations from pre-industrial equilibrium: surface temperature anomaly $T(t)$, deep ocean temperature anomaly $O(t)$, and atmospheric CO_2 concentration $C(t)$ in parts per million.

The surface temperature evolution equation (1) governs energy balance for the coupled surface-atmosphere system, where temperature change rates depend on net energy imbalance scaled by effective heat capacity C_T. The radiative forcing term $F(C) = \alpha \log_2(C/C_0)$ captures the logarithmic relationship between atmospheric CO_2 and absorbed longwave radiation, reflecting fundamental radiative physics where each CO_2 doubling produces equivalent warming effects. The climate feedback parameter λ quantifies radiative response to temperature perturbations via enhanced outgoing longwave radiation, providing primary stabilizing negative feedback. The oceanic heat exchange term $\kappa(T - O)$ represents heat transfer between surface and deep ocean reservoirs.

The deep ocean temperature equation (2) describes thermal evolution through heat exchange with the surface system. Oceanic heat capacity C_O significantly exceeds atmospheric equivalent, introducing multi-decadal response timescales that moderate surface temperature changes while storing substantial thermal energy.

The carbon cycle equation (3) represents atmospheric CO_2 mass balance, with anthropogenic emissions $E(t)$ as primary carbon source and natural sinks providing removal mechanisms. The linear uptake term γC captures CO_2 removal processes, while the temperature-dependent term βTC represents biogeochemical feedback where warming reduces carbon sink efficiency.

Our proposed model's predictive capability emerges from nonlinear coupling between energy and carbon components, creating feedback pathways through radiative forcing, thermal coupling, and temperature-carbon interactions that produce complex temporal behaviors including delayed responses and long-term climate commitment.

Neural ODE Training Configurations and Results

This subsection documents complete hyperparameter specifications, architectural search spaces, optimization strategies, and forecasting performance metrics for Neural ODE implementations across three random seed initializations (seeds 90, 91, 92) with comprehensive training dynamics and convergence analysis under 1% noise conditions.

All Neural ODE configurations employed in this study share a common subset of hyperparameters determined through the optimization procedure. The output dimension was uniformly fixed at $d_{\text{out}} = 3$ across all random seed initializations to ensure compatibility with the three-dimensional state space of the underlying dynamical system. The softplus activation function was consistently selected for all hidden layers throughout the network architecture, providing continuously differentiable nonlinear transformations with smooth gradients conducive to stable ODE integration.

The second optimization phase utilized the AdamW optimizer across all configurations with fixed hyperparameters: momentum coefficients $\beta_1 = 0.9$ and $\beta_2 = 0.999$, weight decay parameter $\lambda = 1 \times 10^{-4}$, and numerical stability constant $\epsilon = 1 \times 10^{-8}$. The learning rate was set to $\eta = 1 \times 10^{-5}$ for both optimization phases across all three random seed

realizations, exhibiting robust convergence characteristics at this scale.

The neural network architecture search space encompassed $n_{\text{layers}} \in [2, 8]$ hidden layers, with each layer comprising $n_{\text{units}} \in [8, 128]$ hidden units. The activation function was selected from the candidate set $\{\text{ReLU}, \text{ELU}, \text{Swish}, \text{Softplus}, \text{Tanh}, \text{Sigmoid}\}$. During the first optimization phase employing the Adam optimizer, the learning rate search space spanned $\eta_1 \in [1 \times 10^{-7}, 1 \times 10^{-1}]$ with maximum iteration counts ranging from $N_{\text{iter}}^{(1)} \in [4{,}000, 40{,}000]$. The second optimization phase explored maximum iterations in the range $N_{\text{iter}}^{(2)} \in [2{,}000, 40{,}000]$.

The following tables present the specific hyperparameter configurations and selected values for each random seed initialization, along with the corresponding training dynamics and forecasting performance metrics.

Table 11: Hyperparameter configuration for Neural ODE with seed 90 and 1% noise

Parameter	Selected Value
Hidden layers	4
Hidden units per layer	64
First optimization iterations	20,000
Second optimization iterations	20,000

Table 12: Hyperparameter configuration for Neural ODE with seed 91 and 1% noise

Parameter	Selected Value
Hidden layers	4
Hidden units per layer	64
First optimization iterations	20,000
Second optimization iterations	20,000

Table 13: Hyperparameter configuration for Neural ODE with seed 92 and 1% noise

Parameter	Selected Value
Hidden layers	5
Hidden units per layer	64
First optimization iterations	20,000
Second optimization iterations	20,000

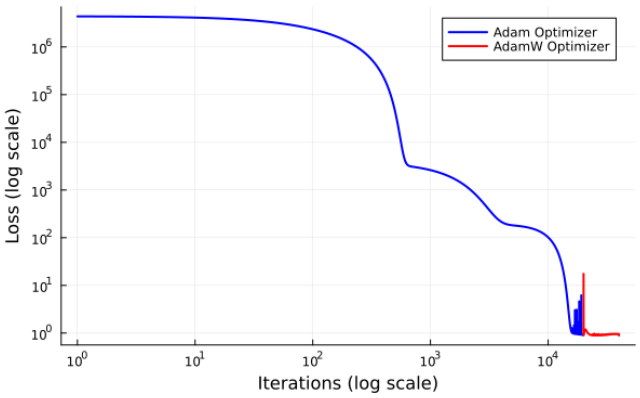

Figure 11: Seed 90 NODE Training Loss Convergence

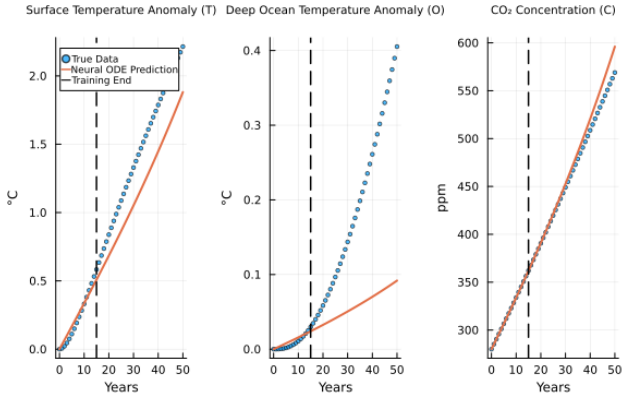

Figure 12: Seed 90 NODE Climate Predictions with 1% Noise

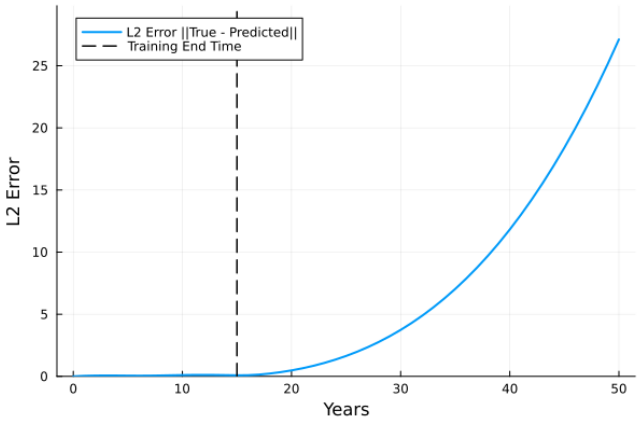

Figure 13: Seed 90 NODE L2 Error Over Time

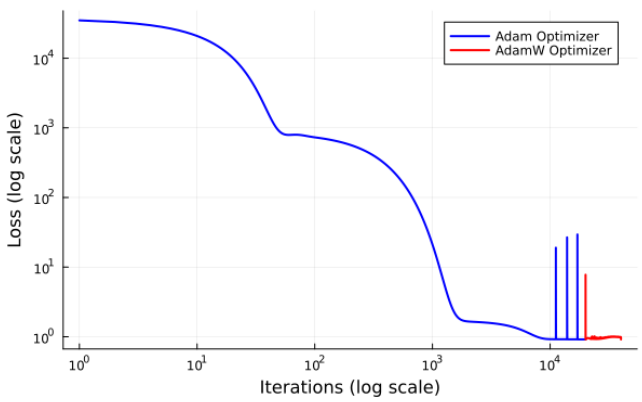

Figure 14: Seed 91 NODE Training Loss Convergence

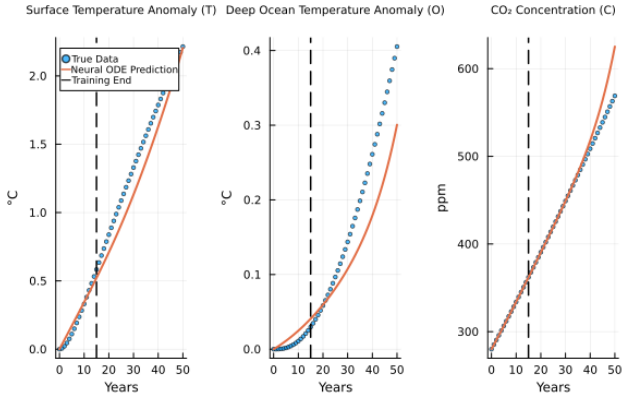

Figure 15: Seed 91 NODE Climate Predictions with 1% Noise

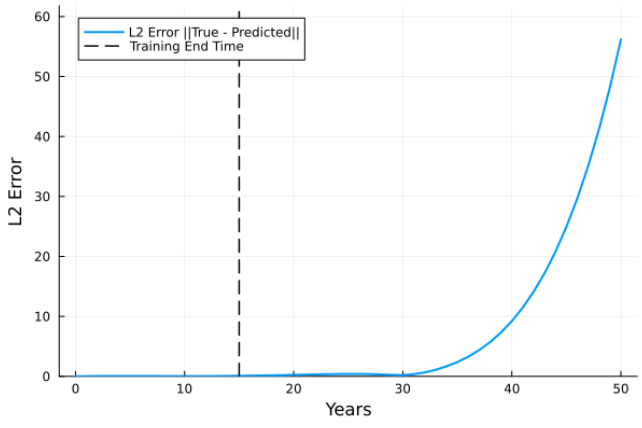

Figure 16: Seed 91 NODE L2 Error Over Time

Table 14: Neural ODE training metrics with random seed 90 and 1% noise

Parameter	Value
Training iterations	40,002
Initial training loss	4.37×10^6
Final training loss	0.893
Loss reduction	99.9999796%

Table 15: Neural ODE training metrics with random seed 91 and 1% noise

Parameter	Value
Training iterations	40,002
Initial training loss	34,927.08
Final training loss	0.916
Loss reduction	99.9973786%

Table 16: Neural ODE training metrics with random seed 92 and 1% noise

Parameter	Value
Training iterations	40,002
Initial training loss	9.85×10^6
Final training loss	0.444
Loss reduction	99.9999955%

Table 17: Neural ODE climate variable predictions at t=50 years with random seed 90 and 1% noise

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	1.881	0.333	15.04
O (°C)	0.405	0.092	0.313	77.33
C (ppm)	568.96	596.08	27.11	4.77

Table 18: Neural ODE climate variable predictions at t=50 years with random seed 91 and 1% noise

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.206	0.009	0.40
O (°C)	0.405	0.301	0.105	25.83
C (ppm)	568.96	625.15	56.19	9.88

Table 19: Neural ODE climate variable predictions at t=50 years with random seed 92 and 1% noise

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	1.729	0.486	21.92
O (°C)	0.405	0.044	0.361	89.09
C (ppm)	568.96	564.00	4.97	0.87

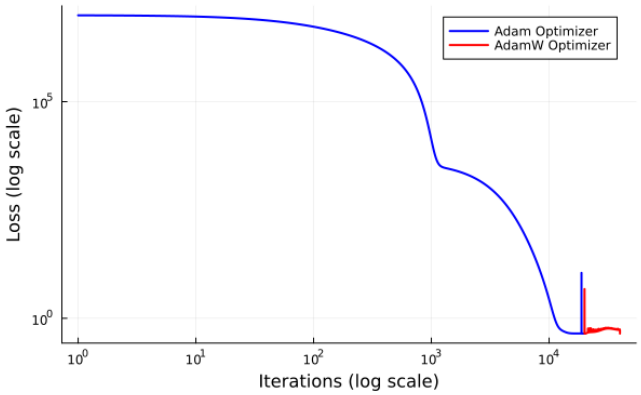

Figure 17: Seed 92 NODE Training Loss Convergence

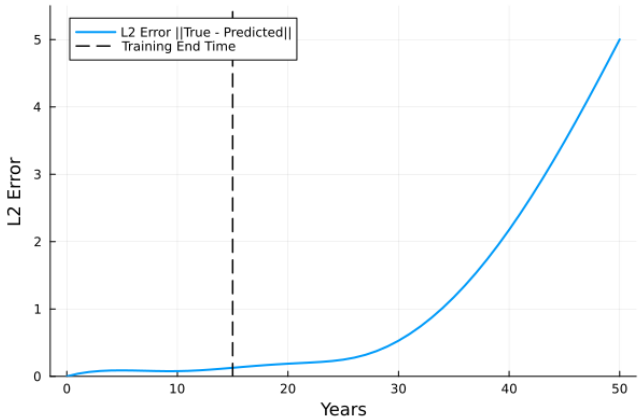

Figure 18: Seed 92 NODE L2 Error Over Time

Universal Differential Equation Training Configurations and Results

This subsection presents exhaustive hyperparameter configurations, neural network architectures, sequential Adam-AdamW optimization specifications, and predictive performance evaluations for UDE frameworks across three random initializations (seeds 90, 91, 92) and six noise perturbation levels (1%, 5%, 10%, 15%, 20%, 25%) with corresponding loss convergence trajectories and climate variable prediction accuracies.

All UDE configurations investigated in this study share a common subset of hyperparameters established through the optimization framework. The output dimension was uniformly constrained to $d_{\text{out}} = 1$ across all random seed initializations, reflecting the architectural design wherein the learned neural network component augments the known physics-based model structure through a single scalar-valued function. The softplus activation function was consistently employed across all hidden layers in each configuration, ensuring continuously differentiable transformations with stable gradient propagation throughout the training procedure. The base architecture consistently incorporated $n_{\text{units}}^{(1)} = 64$ hidden units in the initial layers, with select configurations expanding to $n_{\text{units}}^{(k)} = 128$ units in deeper layers to augment the network's representational capacity.

The neural network architecture search space encompassed $n_{\text{layers}} \in [2, 8]$ hidden layers, with each layer comprising $n_{\text{units}} \in [8, 128]$ hidden units. The activation function was selected from the candidate set $\{\text{ReLU}, \text{ELU}, \text{Swish}, \text{Softplus}, \text{Tanh}, \text{Sigmoid}\}$. During the first optimization phase utilizing the Adam optimizer, the learning rate search space spanned $\eta_1 \in [1 \times 10^{-7}, 1 \times 10^{-1}]$ with maximum iteration counts ranging from $N_{\text{iter}}^{(1)} \in [4{,}000, 40{,}000]$.

For the second optimization phase employing the AdamW optimizer across all three random seed realizations, the hyperparameter search space was systematically explored. Learning rates were sampled from $\eta_2 \in \{1\times10^{-7}, 1\times10^{-6}, 1\times10^{-5}, 1\times10^{-4}, 1\times10^{-3}, 1\times10^{-2}, 1 \times 10^{-1}\}$. Momentum coefficients were selected from $(\beta_1, \beta_2) \in \{(0.9, 0.999), (0.9, 0.9999), (0.99, 0.999), (0.99, 0.9999), (0.999, 0.9999), (0.9999, 0.99999)\}$. Weight decay parameters spanned $\lambda \in \{1 \times 10^{-8}, 1 \times 10^{-7}, 1 \times 10^{-6}, 1 \times 10^{-5}, 1 \times 10^{-4}, 1 \times 10^{-3}\}$. Numerical stability constants were drawn from $\epsilon \in \{1\times10^{-10}, 1\times10^{-9}, 1\times10^{-8}, 1\times10^{-7}, 1\times10^{-6}\}$. Maximum iteration counts ranged from $N_{\text{iter}}^{(2)} \in [2{,}000, 40{,}000]$.

The following tables present the specific hyperparameter configurations and optimal values identified for each random seed initialization, accompanied by comprehensive training dynamics and forecasting performance metrics.

Table 20: Hyperparameter configuration for UDE with random seed 90 and 1% noise

Parameter	Selected Value
Hidden layers	5
Hidden units per layer	64, 64, 64, 64, 128
First optimization learning rate	1×10^{-6}
First optimization iterations	10,000
Second optimization learning rate	1×10^{-6}
Second optimization momentum	(0.9999, 0.99999)
Second optimization weight decay	1×10^{-7}
Second optimization iterations	10,000

Table 21: Hyperparameter configuration for UDE with random seed 91 and 1% noise

Parameter	Selected Value
Hidden layers	3
Hidden units per layer	64
First optimization learning rate	1×10^{-3}
First optimization iterations	5,500
Second optimization learning rate	1×10^{-4}
Second optimization momentum	(0.9, 0.999)
Second optimization weight decay	1×10^{-4}
Second optimization iterations	5,500

Table 22: Hyperparameter configuration for UDE with random seed 92 and 1% noise

Parameter	Selected Value
Hidden layers	4
Hidden units per layer	64, 64, 64, 128
First optimization learning rate	1×10^{-6}
First optimization iterations	10,000
Second optimization learning rate	1×10^{-6}
Second optimization momentum	(0.9999, 0.99999)
Second optimization weight decay	1×10^{-7}
Second optimization iterations	10,000

Table 23: UDE training metrics with random seed 90 and 1% noise

Parameter	Value
Training iterations	20,002
Initial training loss	5,772.58
Final training loss	0.892
Loss reduction	99.9845499%

Table 24: UDE training metrics with random seed 91 and 1% noise

Parameter	Value
Training iterations	11,002
Initial training loss	4,992.94
Final training loss	0.885
Loss reduction	99.9822724%

Table 25: UDE training metrics with random seed 92 and 1% noise

Parameter	Value
Training iterations	20,002
Initial training loss	93.97
Final training loss	0.402
Loss reduction	99.5726506%

Table 26: Training summary for UDE with 5% noise (Random Seed 91)

Metric	Value
Total Training Iterations	11002
Initial Training Loss	5082.45
Final Training Loss	22.13
Loss Reduction (%)	99.56

Table 27: Training summary for UDE with 10% noise (Random Seed 91)

Metric	Value
Total Training Iterations	11002
Initial Training Loss	5234.85
Final Training Loss	88.54
Loss Reduction (%)	98.31

Table 28: Training summary for UDE with 15% noise (Random Seed 91)

Metric	Value
Total Training Iterations	11002
Initial Training Loss	5432.26
Final Training Loss	199.21
Loss Reduction (%)	96.33

Table 29: Training summary for UDE with 20% noise (Random Seed 91)

Metric	Value
Total Training Iterations	11002
Initial Training Loss	5674.68
Final Training Loss	354.14
Loss Reduction (%)	93.76

Table 30: Training summary for UDE with 25% noise (Random Seed 91)

Metric	Value
Total Training Iterations	11002
Initial Training Loss	5962.11
Final Training Loss	533.59
Loss Reduction (%)	91.05

Table 31: UDE climate variable predictions at t=50 years with random seed 90 and 1% noise

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.210	0.004	0.20
O (°C)	0.405	0.405	0.0004	0.11
C (ppm)	568.96	568.01	0.95	0.17

Table 32: UDE climate variable predictions at t=50 years with random seed 91 and 1% noise

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.214	0.0008	0.04
O (°C)	0.405	0.405	0.0003	0.08
C (ppm)	568.96	569.01	0.04	0.007

Table 33: UDE climate variable predictions at t=50 years with random seed 92 and 1% noise

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.219	0.004	0.19
O (°C)	0.405	0.406	0.0006	0.16
C (ppm)	568.96	569.80	0.83	0.15

Table 34: Final variable comparison for UDE with 5% noise (Random Seed 91)

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.208	0.006	0.29
O (°C)	0.405	0.404	0.001	0.33
C (ppm)	568.96	568.00	0.96	0.17

Table 35: Final variable comparison for UDE with 10% noise (Random Seed 91)

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.203	0.012	0.52
O (°C)	0.405	0.403	0.002	0.59
C (ppm)	568.96	567.10	1.86	0.33

Table 36: Final variable comparison for UDE with 15% noise (Random Seed 91)

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.209	0.006	0.25
O (°C)	0.405	0.403	0.002	0.56
C (ppm)	568.96	568.51	0.46	0.08

Table 37: Final variable comparison for UDE with 20% noise (Random Seed 91)

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	2.205	0.009	0.42
O (°C)	0.405	0.402	0.003	0.74
C (ppm)	568.96	567.75	1.22	0.21

Table 38: Final variable comparison for UDE with 25% noise (Random Seed 91)

Variable	True	Predicted	Abs. Error	Error (%)
T (°C)	2.214	1.101	1.113	50.27
O (°C)	0.405	0.271	0.135	33.23
C (ppm)	568.96	386.60	182.36	32.05

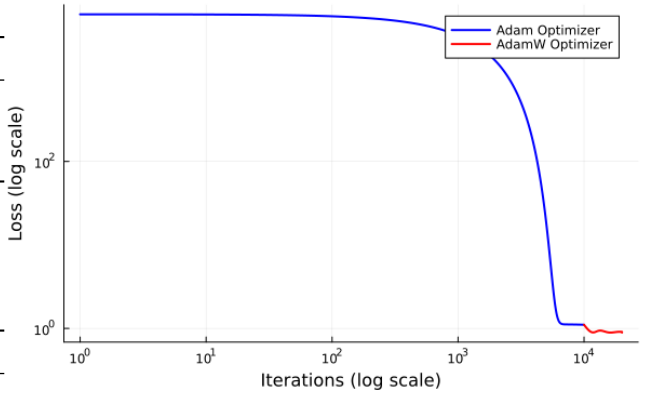

Figure 19: UDE training loss convergence for seed 90 with β learning.

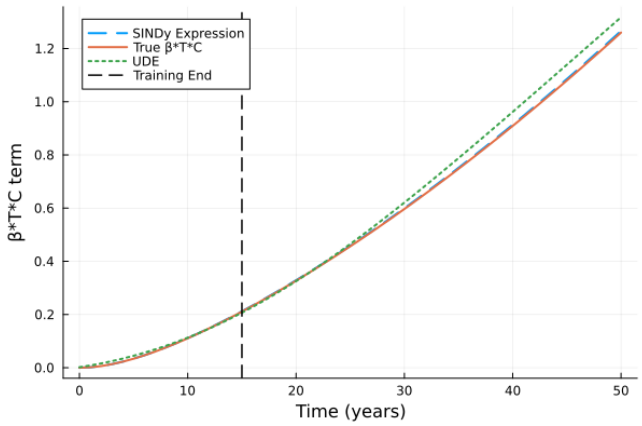

Figure 20: Full timeline comparison of β*T*C term recovery (seed 90).

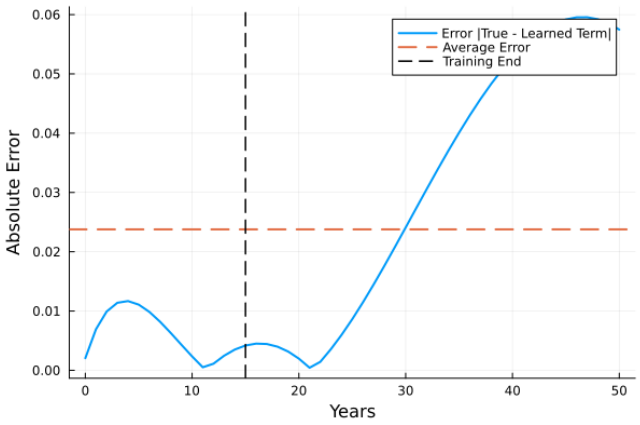

Figure 23: Absolute error in β*T*C term learning (seed 90).

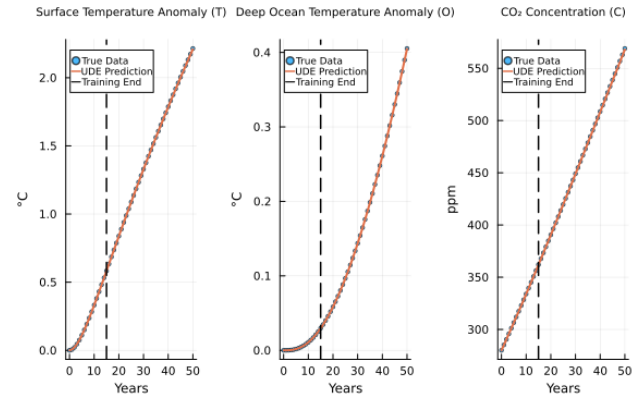

Figure 21: UDE predictions for T, O, C variables (seed 90).

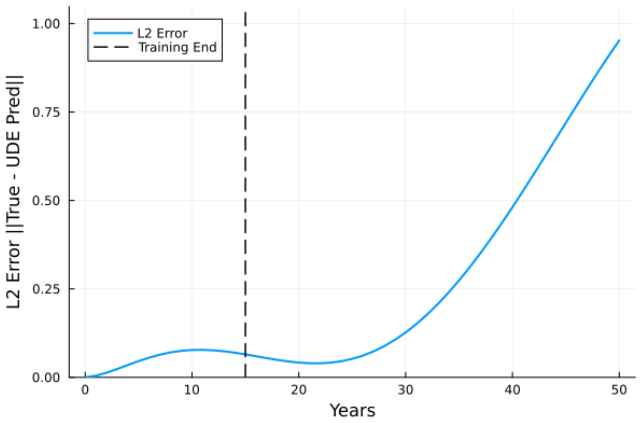

Figure 24: L2 norm prediction error over time (seed 90).

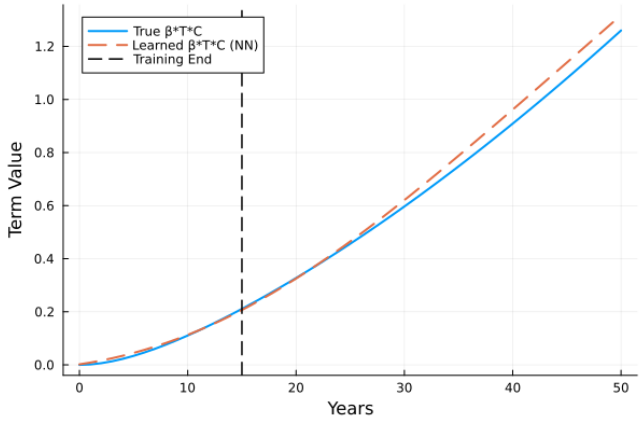

Figure 22: Learned vs true β*T*C term comparison (seed 90).

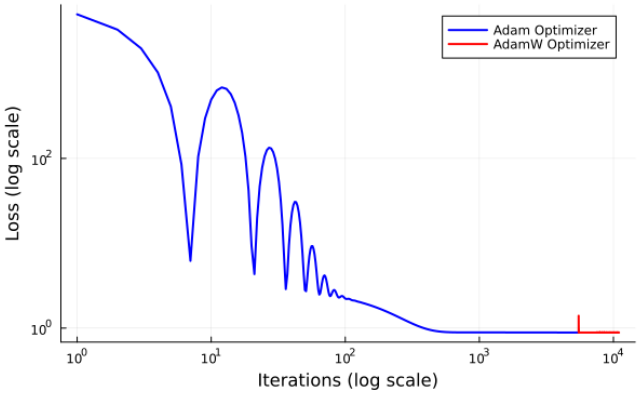

Figure 25: UDE training loss convergence for seed 91 with β learning.

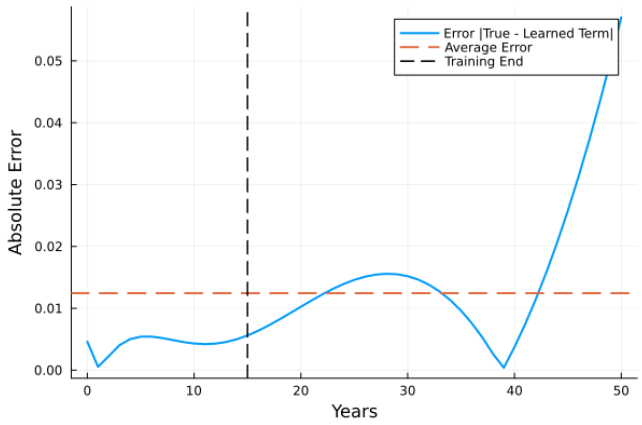

Figure 26: Absolute error in β*T*C term learning (seed 91).

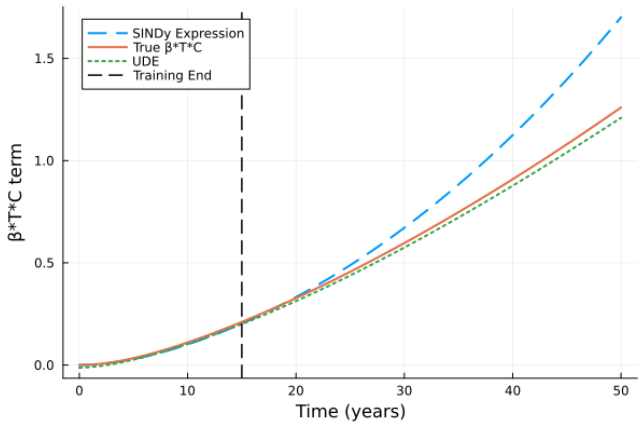

Figure 29: Full timeline comparison of β*T*C term recovery (seed 92).

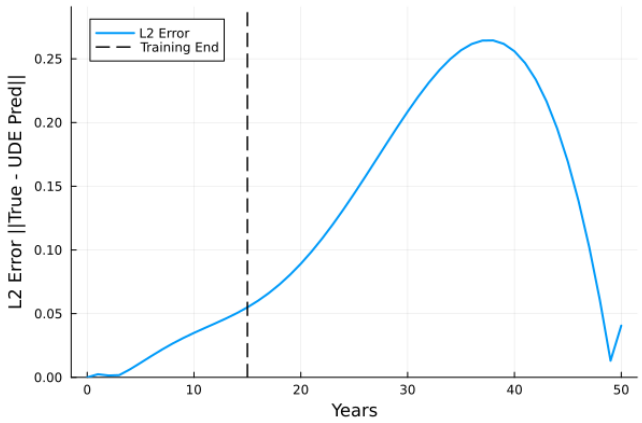

Figure 27: L2 norm prediction error over time (seed 91).

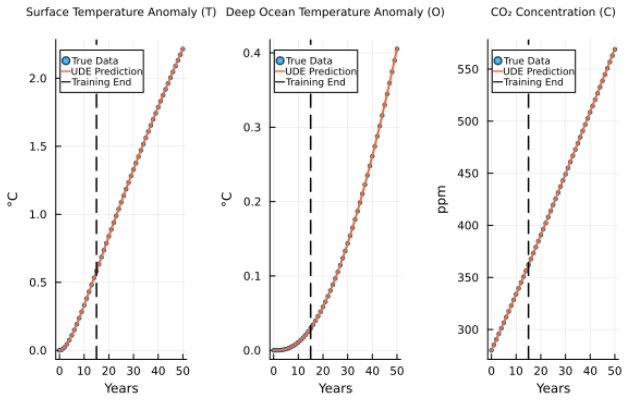

Figure 30: UDE predictions for T, O, C variables (seed 92).

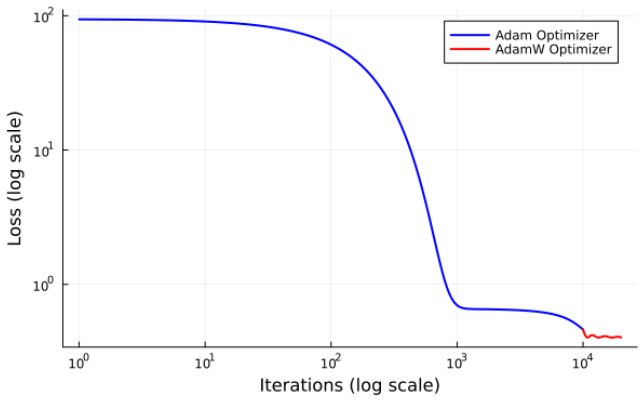

Figure 28: UDE training loss convergence for seed 92 with β learning.

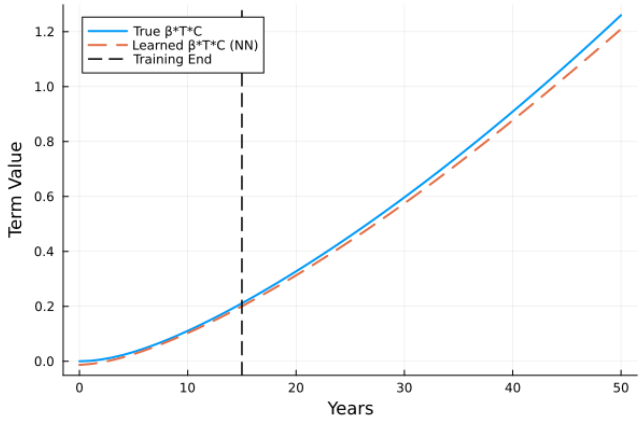

Figure 31: Learned vs true β*T*C term comparison (seed 92).

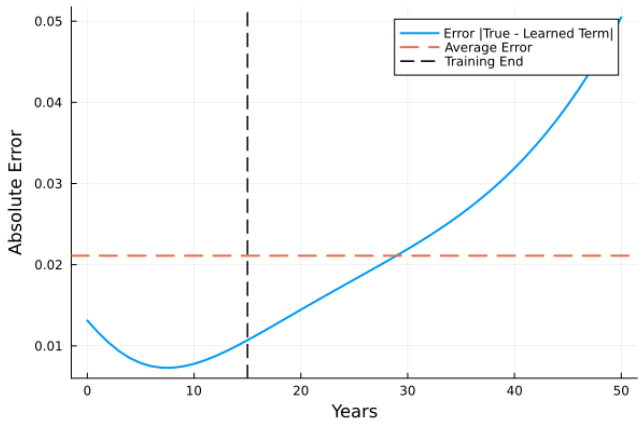

Figure 32: Absolute error in β*T*C term learning (seed 92).

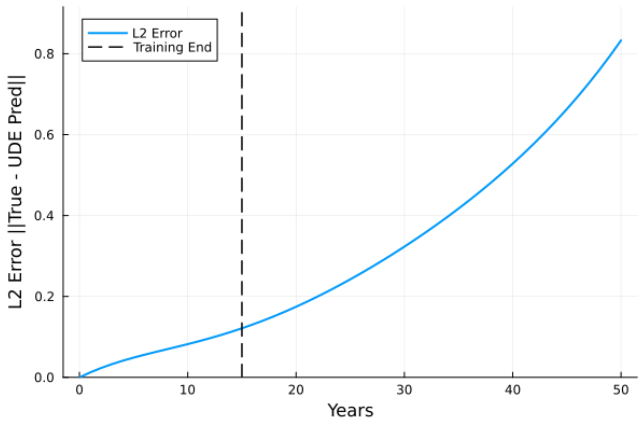

Figure 33: L2 norm prediction error over time (seed 92).

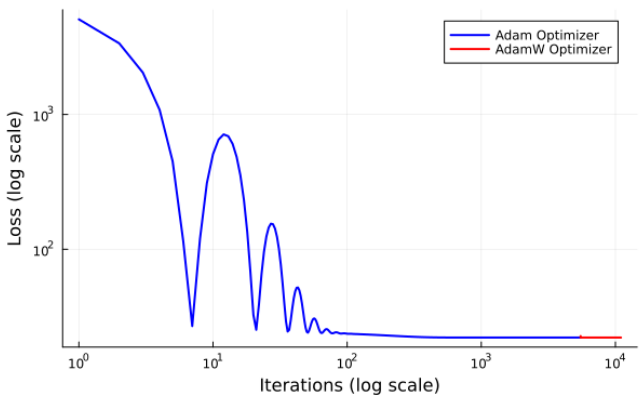

Figure 34: UDE training loss convergence for seed 91 with 5% noise and β learning.

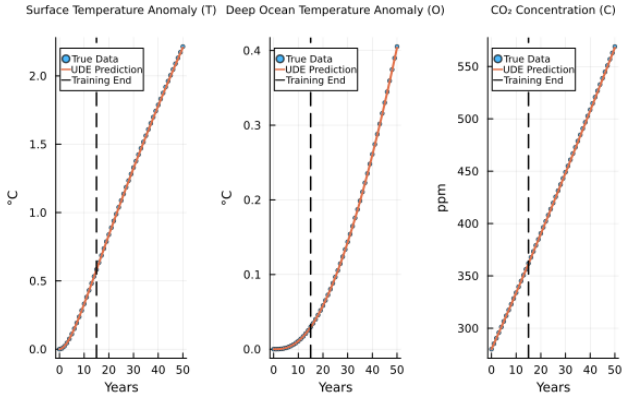

Figure 35: UDE predictions for T, O, C variables with 5% noise (seed 91).

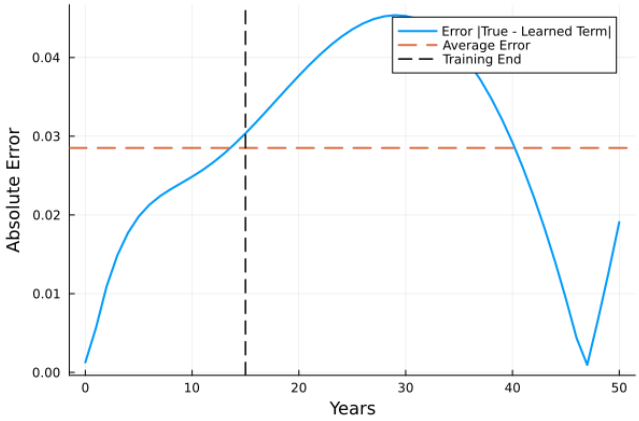

Figure 36: Absolute error in β*T*C term learning with 5% noise (seed 91).

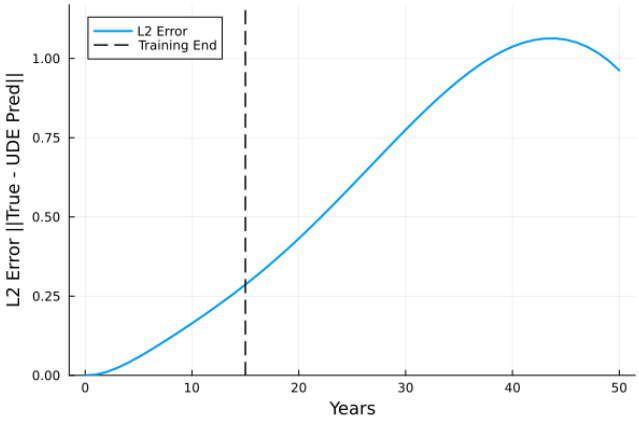

Figure 37: L2 norm prediction error over time with 5% noise (seed 91).

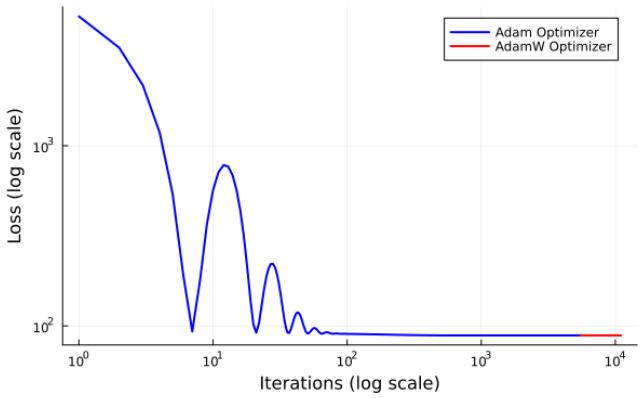

Figure 38: UDE training loss convergence for seed 91 with 10% noise and β learning.

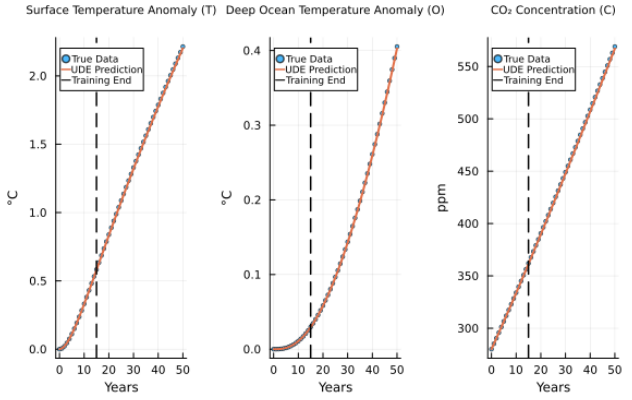

Figure 39: UDE predictions for T, O, C variables with 10% noise (seed 91).

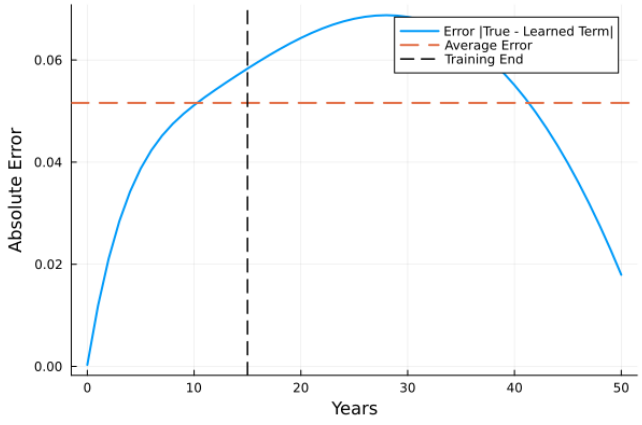

Figure 40: Absolute error in β*T*C term learning with 10% noise (seed 91).

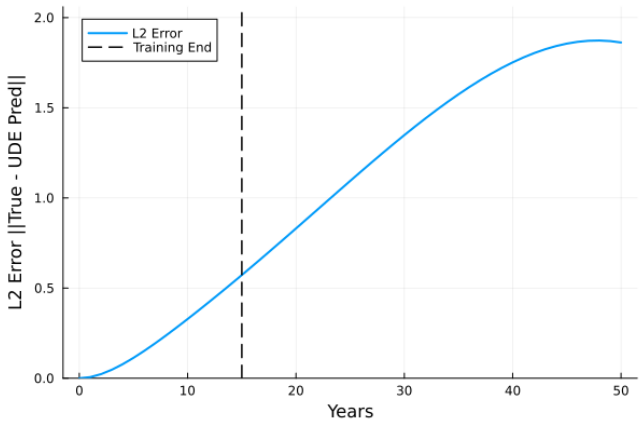

Figure 41: L2 norm prediction error over time with 10% noise (seed 91).

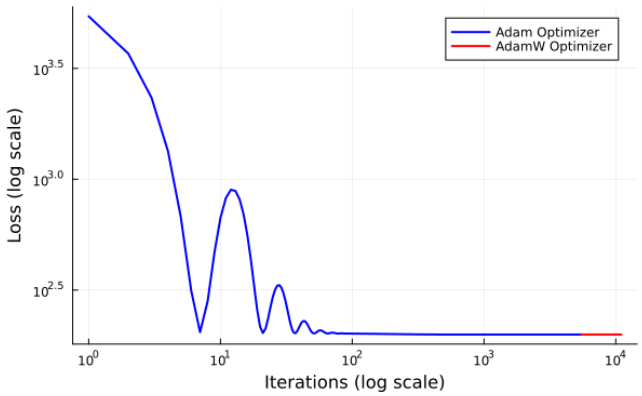

Figure 42: UDE training loss convergence for seed 91 with 15% noise and β learning.

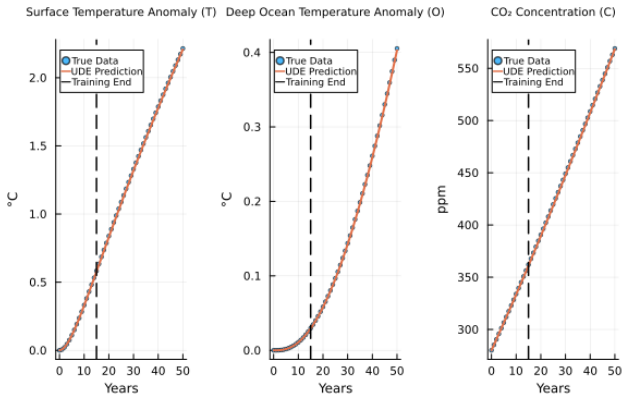

Figure 43: UDE predictions for T, O, C variables with 15% noise (seed 91).

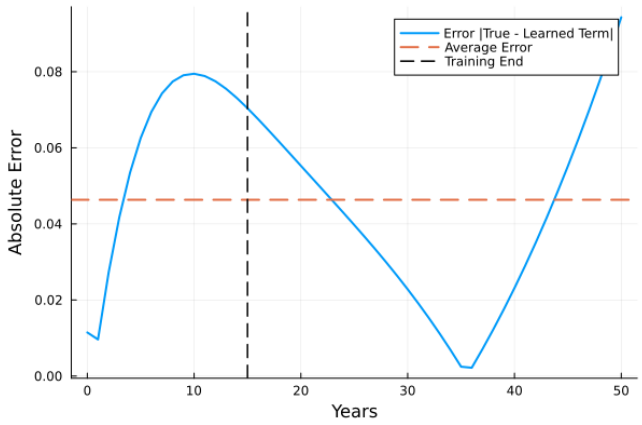

Figure 44: Absolute error in β*T*C term learning with 15% noise (seed 91).

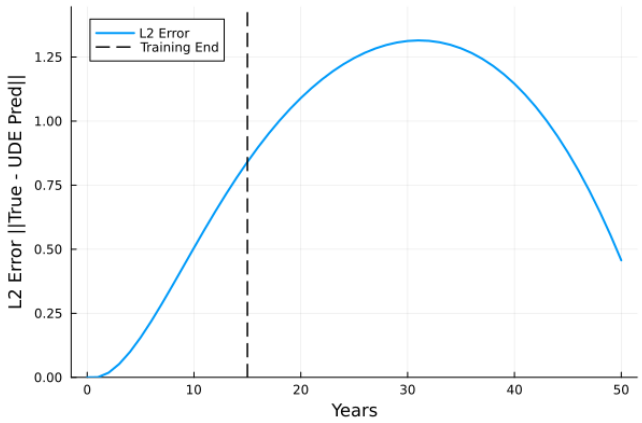

Figure 45: L2 norm prediction error over time with 15% noise (seed 91).

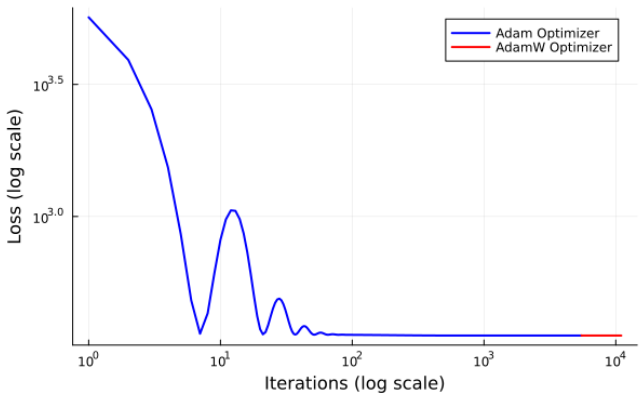

Figure 46: UDE training loss convergence for seed 91 with 20% noise and β learning.

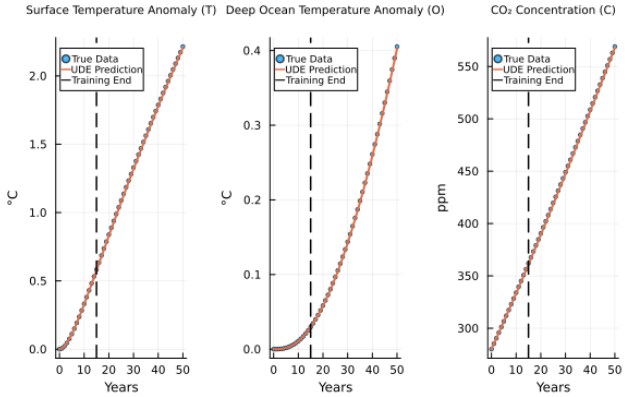

Figure 47: UDE predictions for T, O, C variables with 20% noise (seed 91).

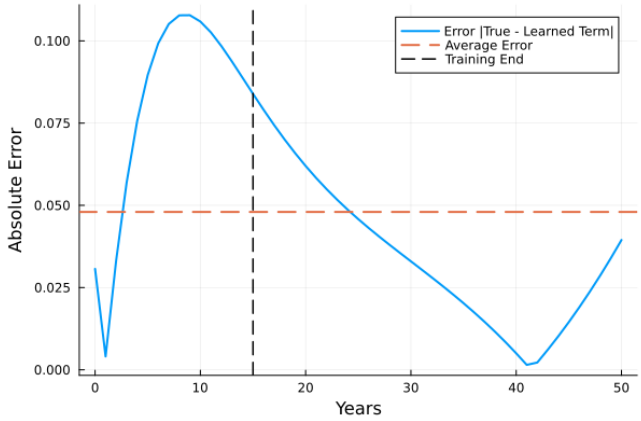

Figure 48: Absolute error in β*T*C term learning with 20% noise (seed 91).

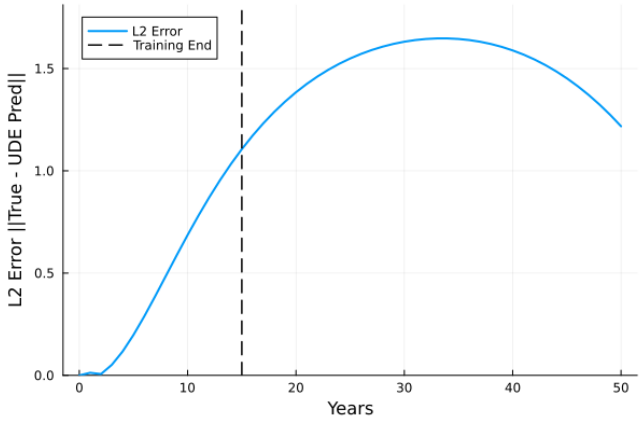

Figure 49: L2 norm prediction error over time with 20% noise (seed 91).

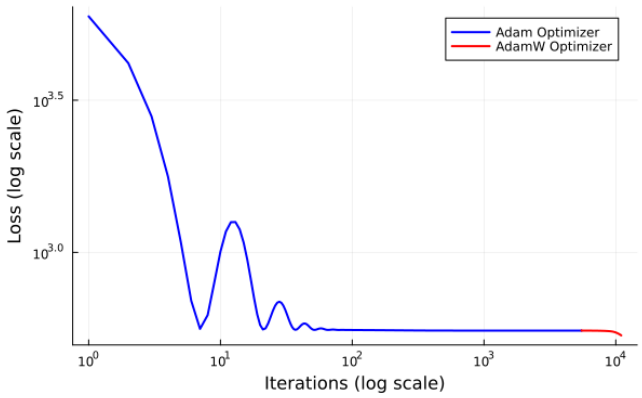

Figure 50: UDE training loss convergence for seed 91 with 25% noise and β learning.

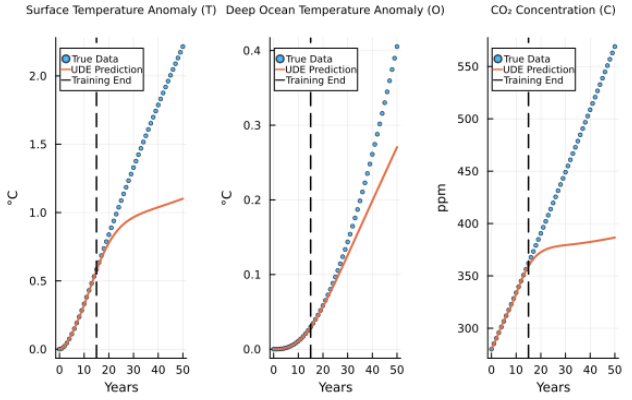

Figure 51: UDE predictions for T, O, C variables with 25% noise (seed 91).

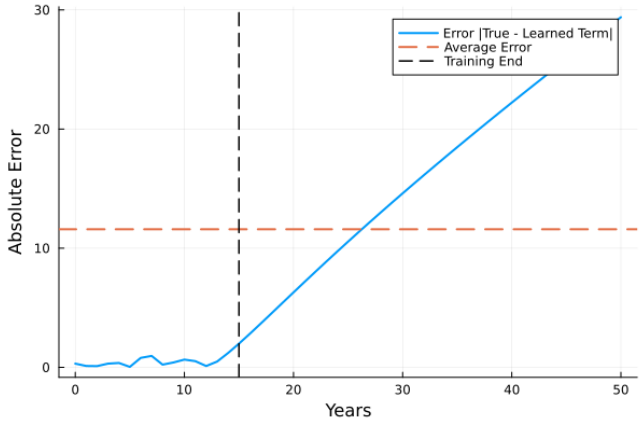

Figure 52: Absolute error in β*T*C term learning with 25% noise (seed 91).

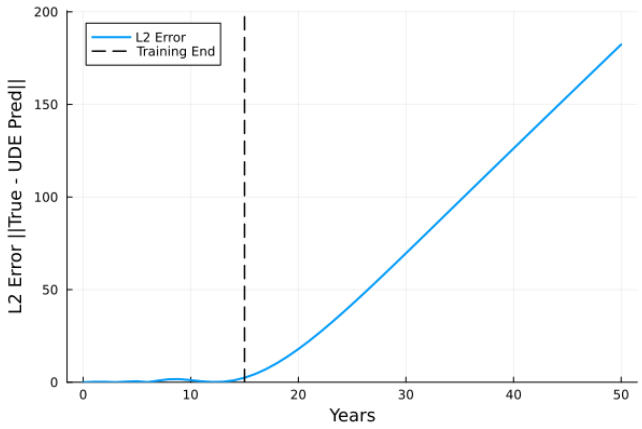

Figure 53: L2 norm prediction error over time with 25% noise (seed 91).

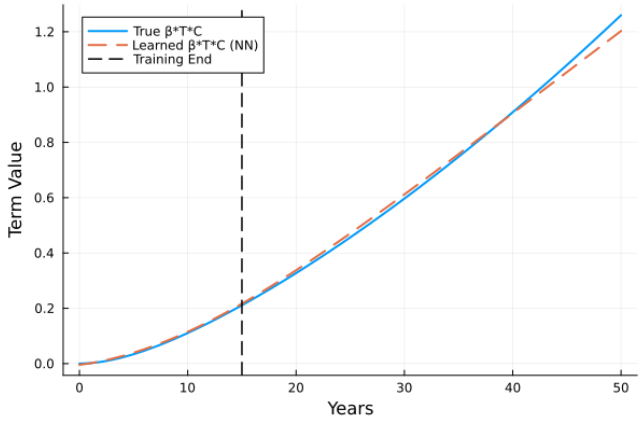

Figure 54: Learned vs true β*T*C term comparison (seed 91).

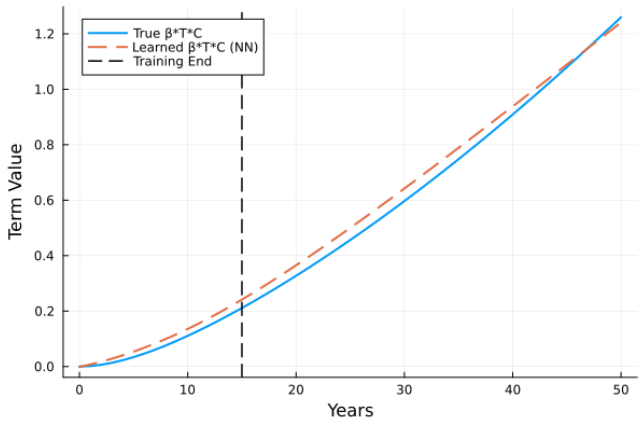

Figure 55: Learned vs true β*T*C term comparison with 5% noise (seed 91).

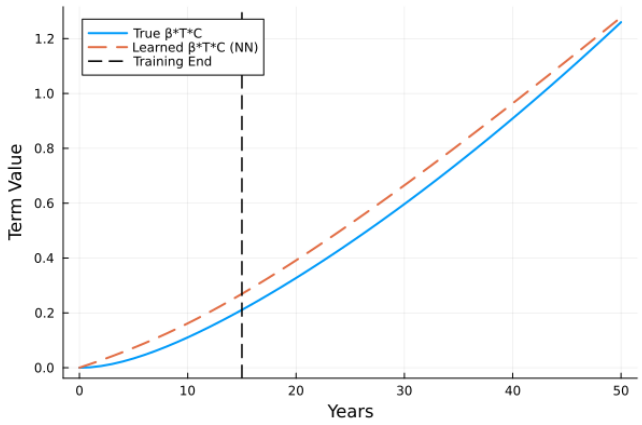

Figure 56: Learned vs true β*T*C term comparison with 10% noise (seed 91).

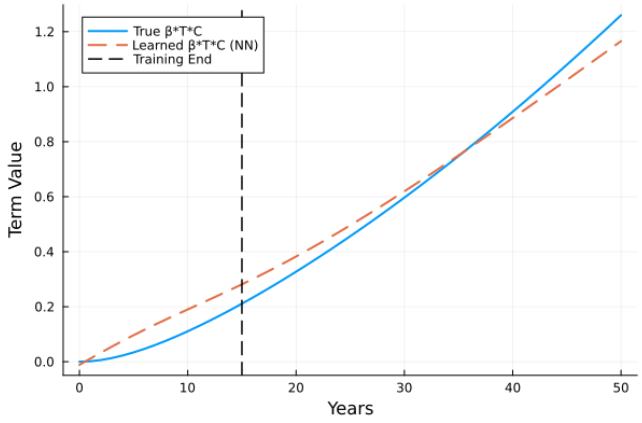

Figure 57: Learned vs true β*T*C term comparison with 15% noise (seed 91).

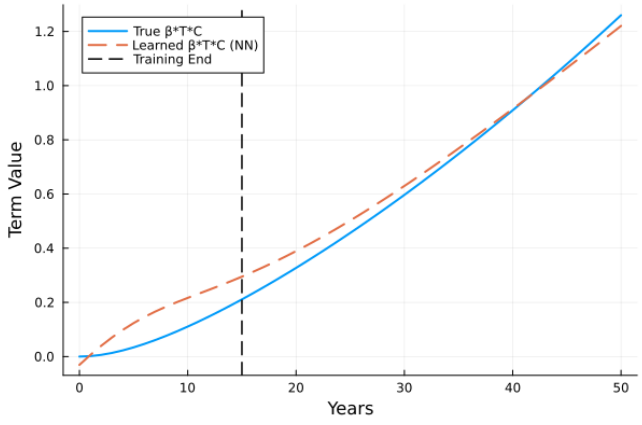

Figure 58: Learned vs true β*T*C term comparison with 20% noise (seed 91).

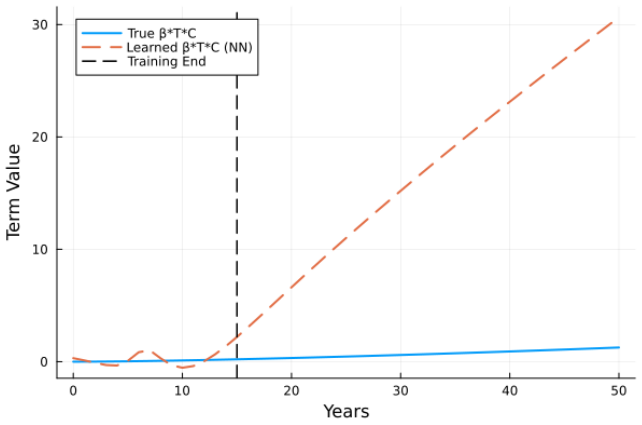

Figure 59: Learned vs true β*T*C term comparison with 25% noise (seed 91).

Non-Negative Least Squares Algorithm

This subsection explicates the mathematical formulation and computational implementation of block principal pivoting methodology for constrained sparse regression, detailing how active-set partitioning and Karush-Kuhn-Tucker optimality enforcement ensure physically interpretable coefficient recovery by systematically eliminating negative values during symbolic expression extraction via SINDy.

The NonNegLeastSquares.jl package implements multiple algorithmic variants for solving the constrained optimization problem $\min_{\mathbf{x}\geq 0} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$, which is fundamental to sparse symbolic regression tasks. The package provides four primary solvers: the classical Lawson-Hanson active-set method, the Fast NNLS variant by Bro and De Jong, the block principal pivoting method by Kim and Park, and an alternating direction method of multipliers approach. The default algorithm employs the block principal pivoting strategy, which partitions variables into passive and active sets corresponding to free and zero-constrained components respectively. This method reformulates the problem as a linear complementarity problem, iteratively identifying which constraints are active by solving unconstrained least squares subproblems on progressively refined variable partitions. The algorithm maintains feasibility through strategic exchanges of multiple variables between working sets, enabling rapid convergence to the Karush-Kuhn-Tucker optimality conditions. Unlike traditional single-variable pivoting schemes that exhibit slow convergence on high-dimensional problems, the block pivoting approach accelerates convergence by simultaneously updating multiple coefficients when favorable, with a conservative fallback mechanism that exchanges individual variables when block updates fail to reduce infeasibility. The computational core involves repeated solutions of reduced normal equations $\mathbf{A}_{\mathcal{P}}^{\top}\mathbf{A}_{\mathcal{P}}\mathbf{x}_{\mathcal{P}} = \mathbf{A}_{\mathcal{P}}^{\top}\mathbf{b}$, where $\mathcal{P}$ denotes the passive set, coupled with gradient evaluations $\mathbf{w} = \mathbf{A}^{\top}(\mathbf{A}\mathbf{x} - \mathbf{b})$ to identify constraint violations. This methodology ensures sparse, physically meaningful coefficient recovery in data-driven discovery frameworks, as negative coefficients—which may be mathematically optimal but physically implausible—are

systematically eliminated through the non-negativity constraint enforcement, thereby yielding interpretable governing equations consistent with domain-specific physical principles.

Table 39: SINDy Recovery of β Coefficient Across Noise Levels (Random Seed 91)

Noise Level (%)	True	Predicted	Abs. Error
1	0.001	0.00103449	0.00003449
5	0.001	0.00119	0.00019
10	0.001	0.00138	0.00038
15	0.001	0.00153	0.00053
20	0.001	0.00169	0.00069
25	0.001	0.00000	0.00100

Table 40: Symbolic regression results for SINDy extraction (Random Seed 91, 1% noise)

Metric	Value	Status
Learned β	0.001034	–
True β	0.001000	–
Relative Error (%)	3.45	Excellent
RMSE	0.002657	–
R^2	0.9985	–

Table 41: Symbolic regression results for SINDy extraction (Random Seed 92, 1% noise)

Metric	Value	Status
Learned β	0.000283	–
True β	0.001000	–
Relative Error (%)	71.70	Poor
RMSE	0.005600	–
R^2	0.9933	–

Table 42: Symbolic regression results for SINDy extraction (Random Seed 90, 1% noise)

Metric	Value	Status
Learned β	0.001005	–
True β	0.001000	–
Relative Error (%)	0.50	Excellent
RMSE	0.007073	–
R^2	0.9876	–

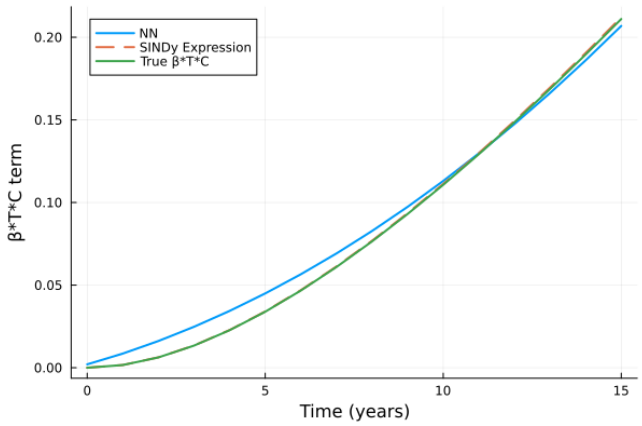

Figure 60: Training period comparison: NN vs SINDy for β*T*C (seed 90).

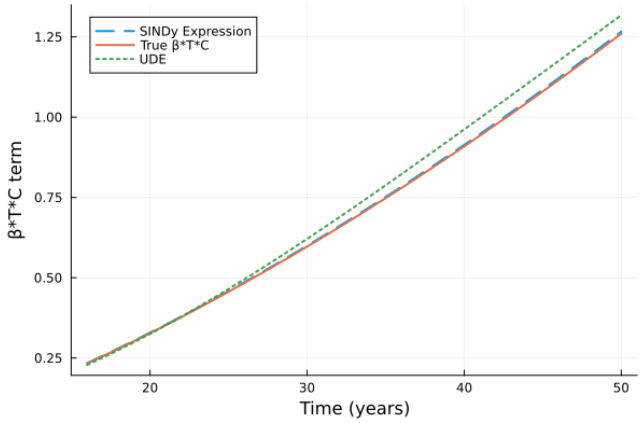

Figure 61: Extrapolation period: SINDy vs true β*T*C term (seed 90).

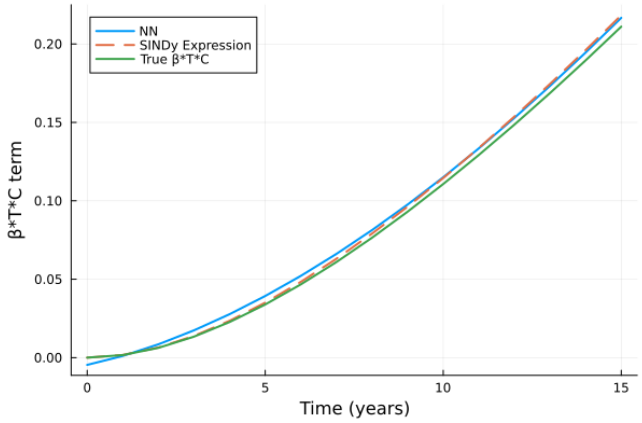

Figure 62: Training period comparison: NN vs SINDy for β*T*C (seed 91).

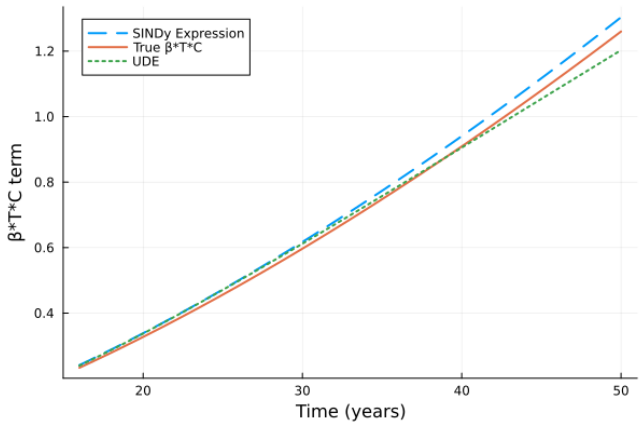

Figure 63: Extrapolation period: SINDy vs true β*T*C term (seed 91).

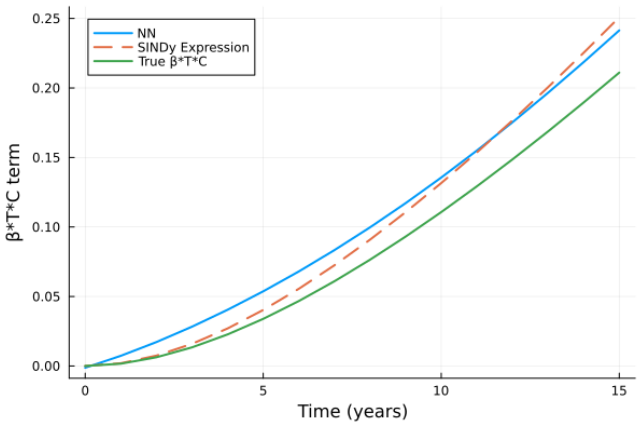

Figure 66: Training period comparison: NN vs SINDy for β*T*C with 5% noise (seed 91).

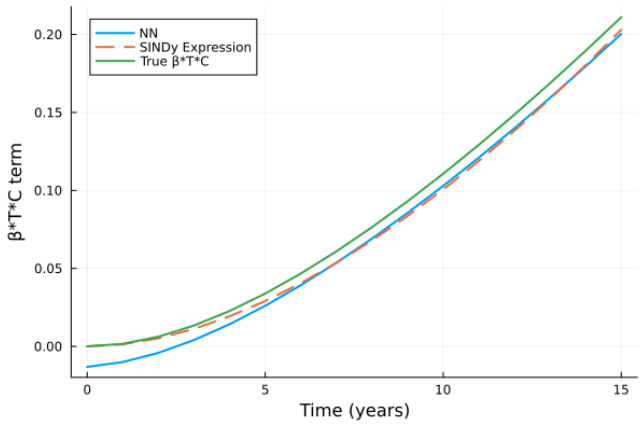

Figure 64: Training period comparison: NN vs SINDy for β*T*C (seed 92).

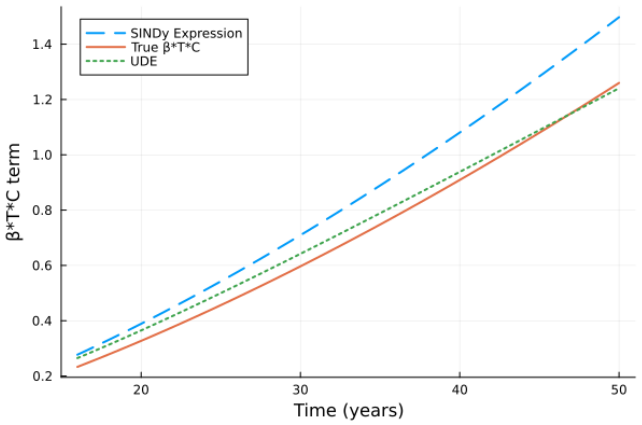

Figure 67: Extrapolation period: SINDy vs true β*T*C term with 5% noise (seed 91).

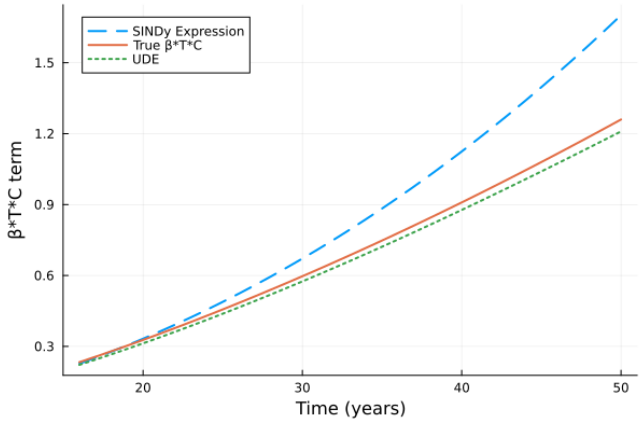

Figure 65: Extrapolation period: SINDy vs true β*T*C term (seed 92).

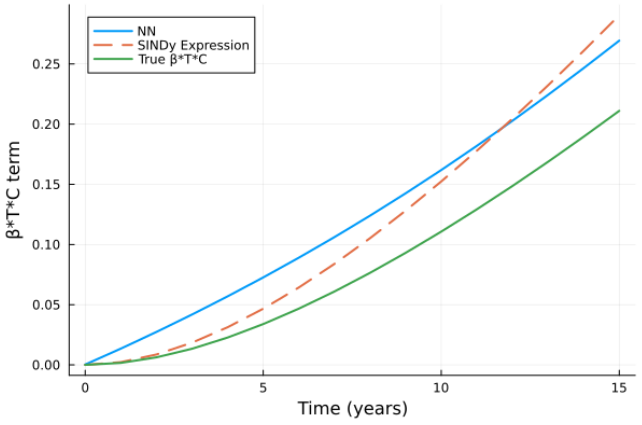

Figure 68: Training period comparison: NN vs SINDy for β*T*C with 10% noise (seed 91).

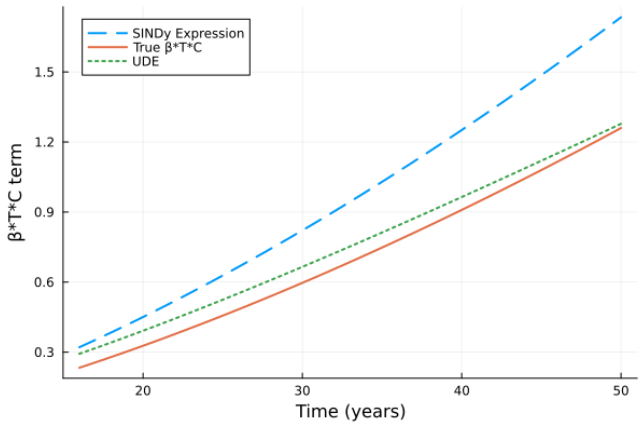

Figure 69: Extrapolation period: SINDy vs true β*T*C term with 10% noise (seed 91).

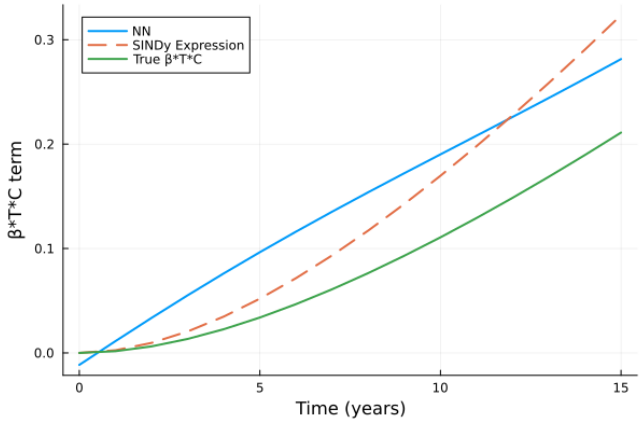

Figure 70: Training period comparison: NN vs SINDy for β*T*C with 15% noise (seed 91).

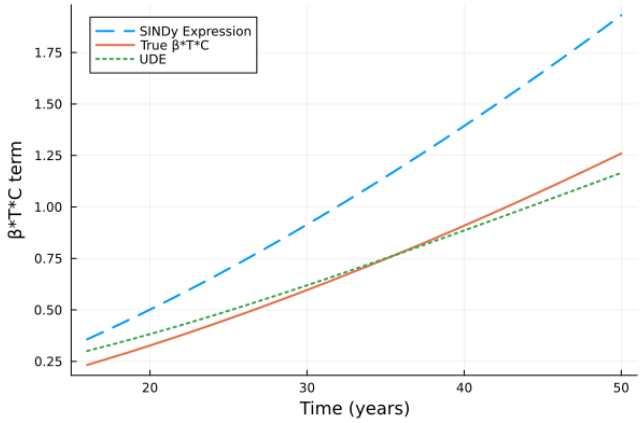

Figure 71: Extrapolation period: SINDy vs true β*T*C term with 15% noise (seed 91).

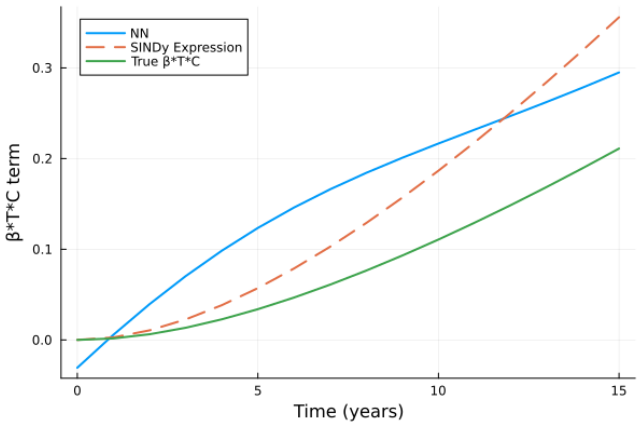

Figure 72: Training period comparison: NN vs SINDy for β*T*C with 20% noise (seed 91).

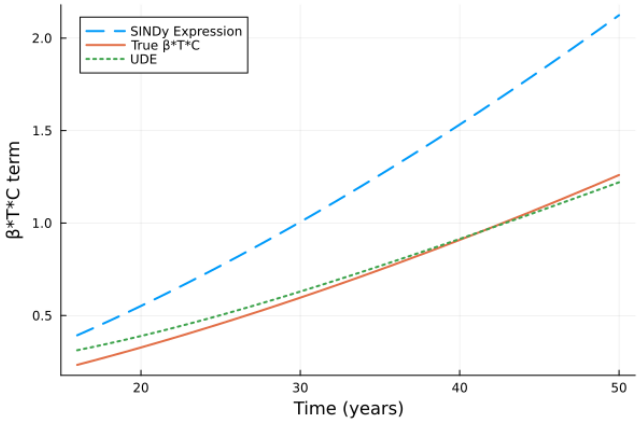

Figure 73: Extrapolation period: SINDy vs true β*T*C term with 20% noise (seed 91).

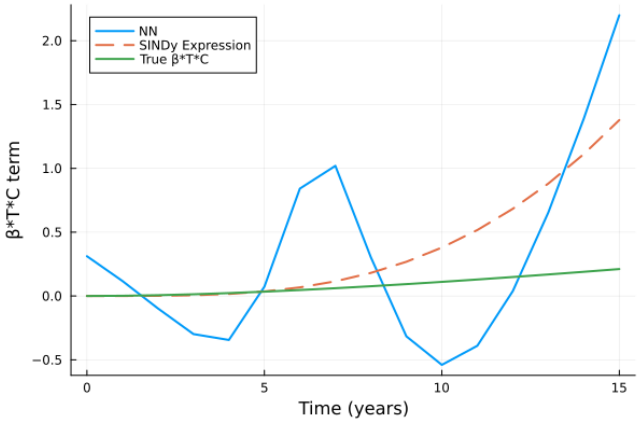

Figure 74: Training period comparison: NN vs SINDy for β*T*C with 25% noise (seed 91).

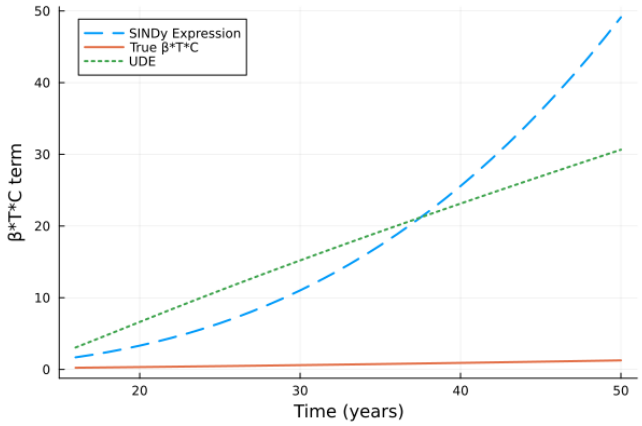

Figure 75: Extrapolation period: SINDy vs true β*T*C term with 25% noise (seed 91).

ARIMA Model Implementation Details

This subsection contains comprehensive hyperparameter optimization results via exhaustive grid search across 48 candidate specifications, detailed AICc criterion evaluations, maximum likelihood estimation procedures, and complete forecasting performance metrics including RMSE, percentage errors, and computational resource comparisons for univariate statistical baseline models.

The Autoregressive Integrated Moving Average (ARIMA) model represents a fundamental statistical framework for time series forecasting, systematically decomposing temporal dynamics into three distinct components. The autoregressive (AR) component captures the dependent relationship between current observations and lagged values, while the moving average (MA) component models the linear combination of past forecast errors, and the integrated (I) component addresses non-stationarity through differencing operations. The model is formally denoted as ARIMA(p,d,q), where p represents the autoregressive order, d specifies the degree of differencing required to achieve stationarity, and q indicates the moving average order.

Our implementation employs the Box-Jenkins methodology, a systematic iterative approach comprising model identification, parameter estimation, and diagnostic checking. For each climate variable—surface temperature anomaly (T), deep ocean temperature anomaly (O), and atmospheric CO_2 concentration (C)—we construct independent univariate ARIMA models using the training period observations ($t \in [0, 15]$ years). Model selection is conducted through a grid search across candidate parameter combinations, with optimal orders determined by minimizing the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which balance goodness-of-fit against model complexity. The StateSpaceModels.jl package in Julia facilitates maximum likelihood estimation of model coefficients.

The fitted ARIMA models function as temporal filters, extracting deterministic signal patterns from noisy observations and subsequently extrapolating these patterns into the forecast horizon. For the extrapolation period ($t \in (15, 50]$ years), each model generates point forecasts through recursive application of the estimated autoregressive and moving average operators to the differenced series, with predictions subsequently transformed back to the original scale through cumulative summation. This purely statistical approach provides a baseline comparison against physics-informed neural architectures, explicitly testing whether data-driven temporal correlation structures can adequately capture the underlying physical climate dynamics without incorporating domain-specific mechanistic knowledge.

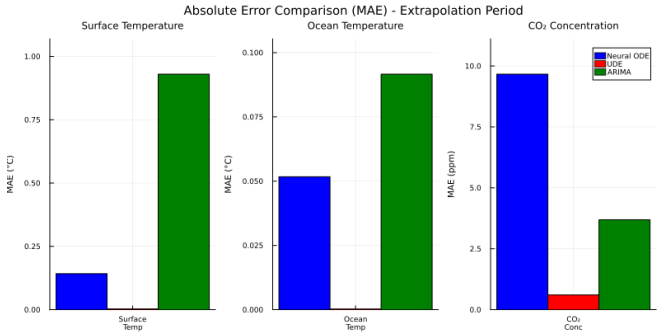

Figure 76: Absolute prediction errors for Neural ODE, UDE, and ARIMA models across all three climate variables during extrapolation period.

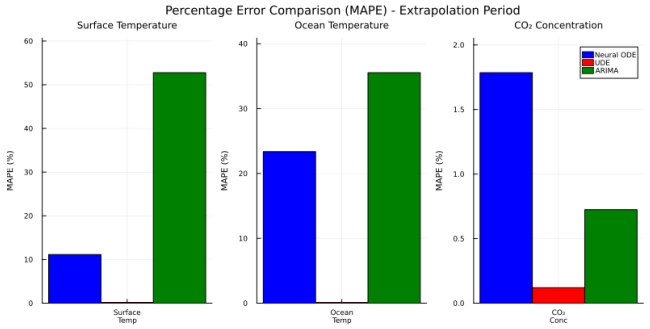

Figure 77: Percentage prediction errors for Neural ODE, UDE, and ARIMA models across all three climate variables during extrapolation period.

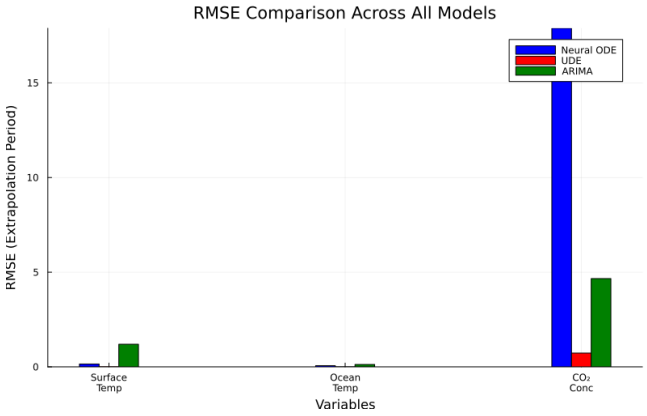

Figure 78: Root Mean Square Error (RMSE) comparison across Neural ODE, UDE, and ARIMA models for all climate variables.

Table 43: Optimal ARIMA Parameters Selected via Grid Search with AICc Criterion

Variable	ARIMA(p,d,q)	AICc
Surface Temperature (T)	(3, 0, 0)	-101.61
Ocean Temperature (O)	(2, 0, 0)	-198.05
CO_2 Concentration (C)	(0, 2, 2)	24.09

Table 44: Computational Resource Usage Across Models

Metric	Neural ODE	UDE	ARIMA
Parameters	12,931	8,577	10
Training Iterations	40,000	11,000	144
Training Time (s)	1,036.99	193.35	7.52

Table 45: Extrapolation Performance: RMSE and R^2 Scores (t = 15 to 50 years)

Variable	Neural ODE	UDE	ARIMA
RMSE			
Surface Temp. (°C)	0.1516	**0.0037**	1.1946
Ocean Temp. (°C)	0.0637	**0.0004**	0.1260
CO_2 Conc. (ppm)	17.8834	**0.7291**	4.6630
R^2 Score			
Surface Temp.	0.8958	**0.9999**	-5.4675
Ocean Temp.	0.6760	**0.9999**	-0.2673
CO_2 Conc.	0.9108	**0.9999**	0.9939

Vector Autoregression Implementation Details

This subsection provides rigorous mathematical formulation of multivariate VAR(p) framework, lag order selection via BIC criterion, ordinary least squares estimation procedures with contemporaneous error correlation structures, stability verification through companion matrix eigenvalue analysis, and recursive multi-step-ahead forecasting performance across all climate variables.

The Vector Autoregressive (VAR) framework extends univariate autoregressive modeling to multivariate time series analysis, enabling the capture of bidirectional interdependencies among multiple climate system variables. In this implementation, a VAR(p) model of order p was specified to forecast the coupled dynamics of surface temperature anomaly (T), deep ocean temperature anomaly (O), and atmospheric CO_2 concentration (C).

The VAR(p) model represents each variable as a linear function of its own lagged values and the lagged values of all other endogenous variables in the system. Mathematically, the trivariate VAR(p) model for our climate system is expressed as:

$$\mathbf{Y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{Y}_{t-i} + \boldsymbol{\varepsilon}_t \tag{9}$$

where $\mathbf{Y}_t = [T_t, O_t, C_t]^\top$ denotes the state vector at time t, $\mathbf{c}$ represents the constant intercept vector, $\mathbf{\Phi}_i$ are (3×3) coefficient matrices capturing lagged influences, and $\boldsymbol{\varepsilon}_t$ denotes the white noise error vector with covariance matrix $\mathbf{\Sigma}$.

The optimal lag order p was determined through systematic evaluation of the Bayesian Information Criterion (BIC), which balances model fit against parametric complexity by penalizing excessive parameterization. The BIC for a VAR model with K variables and lag order p is computed as $\text{BIC}(p) = \log(\det(\hat{\mathbf{\Sigma}}_p)) + \frac{pK^2 \log(T)}{T}$, where $\hat{\mathbf{\Sigma}}_p$ denotes the estimated error covariance matrix and T represents the sample size. BIC was selected over alternative criteria due to its theoretical property of consistently estimating the true lag order, whereas AIC tends toward overparameterization asymptotically. Candidate models spanning $p \in \{1, 2, \ldots, p_{\max}\}$ were estimated, with the configuration minimizing BIC designated as optimal.

Parameter estimation employed ordinary least squares (OLS) regression applied equation-by-equation, which yields consistent and asymptotically efficient estimates when error terms exhibit contemporaneous correlation but no cross-equation restrictions are imposed. The trained VAR(p_{optimal}) model subsequently generated recursive multi-step-ahead forecasts through iterative application of the fitted coefficient matrices, propagating uncertainty through the error covariance structure to project climate trajectories beyond the training horizon.

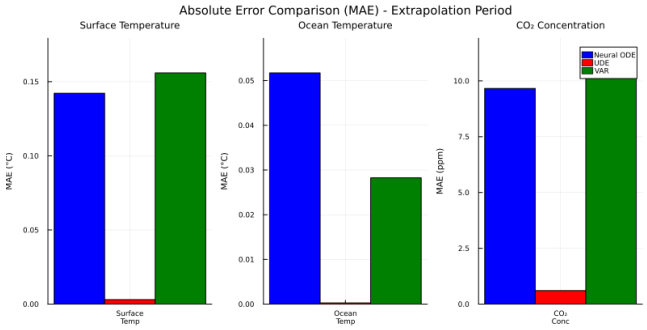

Figure 79: Absolute prediction errors for VAR, Neural ODE, and UDE models with 1% noise (seed 91).

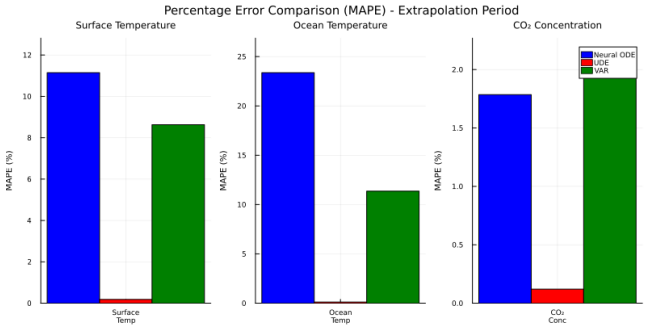

Figure 80: Percentage prediction errors for VAR, Neural ODE, and UDE models with 1% noise (seed 91).

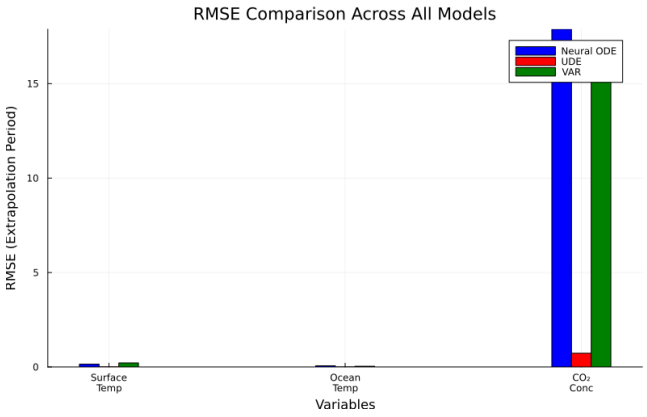

Figure 81: RMSE comparison across all climate variables for VAR, Neural ODE, and UDE models with 1% noise (seed 91).

Table 46: VAR Model Hyperparameter Selection (Random Seed 91, 1% Noise)

Lag Order	AIC	BIC
1	-25.2751	-24.8503
2	-25.6966	-24.8749

Selected: VAR(2) by BIC criterion

Table 47: Computational Resource Usage (Random Seed 91, 1% Noise)

Metric	Neural ODE/UDE	VAR
Parameters	12931/8577	21
Iterations	40000/11000	2
Training Time (s)	907.33/208.81	0.99

Table 48: RMSE for Extrapolation Period ($t = 15$ to 50 years, Seed 91, 1% Noise)

Variable	Neural ODE	UDE	VAR
Surface Temp. (°C)	0.1516	**0.0037**	0.2150
Ocean Temp. (°C)	0.0637	**0.0004**	0.0375
CO_2 Conc. (ppm)	17.8834	**0.7291**	15.2162