Scientific Machine Learning for Symbolic Recovery of Relativistic Effects in Black Hole Orbits

Pothuraju Naveen Yadav
Delhi Technological University
pothurajunaveen_24isy14@dtu.ac.in

Prathamesh Dinesh Joshi
Vizuara AI Labs
prathamesh@vizuara.com

Raj Abhijit Dandekar
Vizuara AI Labs
raj@vizuara.com

Rajat Dandekar
Vizuara AI Labs
rajatdandekar@vizuara.com

Sreedath Panat
Vizuara AI Labs
sreedath@vizuara.com

Dinesh Kumar Vishwakarma
Delhi Technological University
dinesh@dtu.ac.in

Abstract

Simulating relativistic orbital dynamics around Schwarzschild black holes is essential for understanding general relativity and astrophysical phenomena like precession. Traditional numerical solvers face difficulty when dealing with noisy or sparse data, necessitating data-driven approaches. We develop a Scientific Machine Learning (SciML) framework to model orbital trajectories and symbolically recover the relativistic correction term. Neural Ordinary Differential Equations (Neural ODEs) accurately predict inverse radius u, radial velocity v, and precession δ, performing well under ideal conditions with mean absolute errors (MAE) around 3.5×10^{-4} for u on noiseless full-domain data, but degrading sharply when training data is limited to 10%, where MAE rises above 0.026. To address this, we employ Universal Differential Equations (UDEs), which embed a neural network to approximate the correction term $\alpha\,\frac{GM}{c^2}u^3$, achieving precise orbit predictions even with just 10–20% data coverage and maintaining low errors (UDE forecast loss $\approx 2.7\times10^{-4}$) across noise levels up to 35%. Symbolic regression further recovers an analytical expression closely matching the expected correction, with mean symbolic errors below 10^{-7}. We use adjoint-based training to discover solutions efficiently, implemented in Julia with `DiffEqFlux` and `Lux`. Using this method, we successfully integrate machine learning with physical laws, demonstrating robustness to noise and data scarcity. This approach can be expanded for large-scale or detailed astrophysical projects.

1 Introduction

Scientific Machine Learning (SciML) enables data-driven discovery and robust forecasting in physical systems by combining neural models with analytic priors [1, 2, 3, 4, 5]. In this context, Neural Ordinary Differential Equations (Neural ODEs) [6, 7, 8, 9, 10] use deep networks to model unknown dynamics, while Universal Differential Equations (UDEs) [2] augment known physics with a neural correction, offering interpretability and generalization critical for scientific domains such as astrophysics. Despite progress in SciML, systematic benchmarks for relativistic systems remain rare.

Published in the AI for Science workshop (NeurIPS 2025).

We focus on the Schwarzschild orbital equation, which describes the precessing trajectory of a test particle around a compact mass:

$$\frac{du}{d\phi} = v, \qquad \frac{dv}{d\phi} = -u + \frac{GM}{L^2} + \alpha \frac{GM}{c^2} u^3, \qquad \frac{d\delta}{d\phi} = \beta u^2, \tag{1}$$

where $u = 1/r$ is the inverse radius, $v = \frac{du}{d\phi}$ is the radial velocity, δ is the precession angle (the angular shift of the orbit), and $G, M, L, c, \alpha, \beta$ are physical parameters. This Schwarzschild formulation is standard in general relativity [11, 12].

We assess Neural ODEs and UDEs for this system under varying noise levels (0%, 7%, 35%) and data-limited regimes (100%–10% coverage), focusing on forecasting and recovery of the embedded cubic relativistic term. From extensive model variations, we report only the best-performing configurations for each approach to highlight practical capabilities.

2 Methodology

We generate synthetic data by numerically integrating the Schwarzschild system (Eq. 1) using canonical parameter values. Training subsets span 100%–10% of the domain (the remainder is used for forecasting); Gaussian noise is added at tested levels per variable.

2.1 Neural ODE

We treat the right-hand side of the system as a trainable multilayer perceptron (MLP) with two layers and `tanh` activations, with inputs and outputs matching (u, v, δ). Formally, the Neural ODE framework [6] models the dynamics as $\frac{d\mathbf{x}}{d\phi} = f_\theta(\mathbf{x}, \phi)$, where $\mathbf{x} = (u, v, \delta)$ and f_θ is a neural network parameterized by θ. This allows the network to learn all dynamics, including both known and unknown corrective effects, directly from observation. Optimization proceeds by AdamW followed by BFGS for convergence and precision, with strict ODE solver tolerances and regularization for robustness. Multiple architectures, initializations, and schedules were trialed; only the top-performing configuration is reported.

2.2 Universal Differential Equation (UDE)

The UDE formulation [2] retains the analytic structure of the Schwarzschild system, but augments uncertain terms with a neural network. In general, a UDE is written as $\frac{d\mathbf{x}}{d\phi} = f(\mathbf{x}, \phi) + g_\theta(\mathbf{x}, \phi)$, where f represents the known physics and g_θ is a neural correction. For our case, $\frac{dv}{d\phi} = -u + \frac{GM}{L^2} + \alpha \frac{GM}{c^2} u^3 + \hat{g}_\theta(u)$, with $\hat{g}_\theta(u)$ modeled as a compact neural network (two `softplus` layers, output scaled by u^3 and clamped for stability). The loss balances data fit, correction accuracy, and regularization, with adaptive weighting across regimes and the same optimizer protocol as with Neural ODEs.

2.3 Model selection and evaluation

For each method, models are trained on selected noisy/clean subdomains and evaluated by mean absolute errors both in forecasting and on their learned correction term. Only the best results per approach are included.

2.4 Symbolic recovery

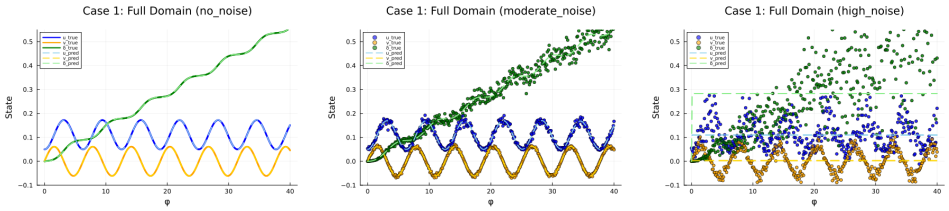
After UDE training, the learned neural correction is symbolically regressed onto a u^3 basis via convex optimization, yielding a physically interpretable coefficient directly comparable to the true analytic correction term $\alpha \frac{GM}{c^2} u^3$. Recovery error quantifies the fidelity of physical knowledge extraction. Symbolic recovery has proven effective for extracting governing laws from data and hybrid models [13, 14, 15].

All experiments use the Julia SciML stack [1, 16, 17, 18, 19].

3 Results

Six cases were considered, corresponding to different percentages of the available dataset and evaluated under three noise conditions: no noise, moderate noise with a 7% standard deviation, and high noise with a 35% standard deviation. The main text presents results for the Neural ODE and UDE models trained on the full dataset, as well as on 80%, 20%, and 10% of the data. The remaining two cases, with 90% and 40% training coverage, are provided in Appendix A.

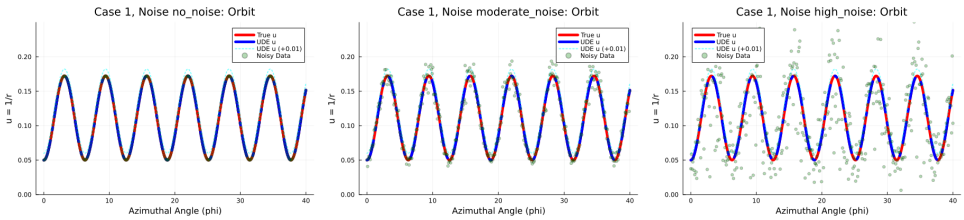
3.1 Case 1: Training on the full orbital domain



(a) no noise (b) moderate noise (c) high noise

Figure 1: Neural ODE results for Case 1 across different noise levels.

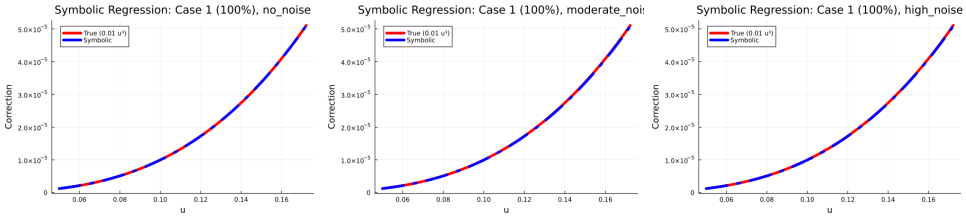
From Figure 1, we can observe that the Neural ODE learns the Schwarzschild orbital dynamics across the entire domain and under various noise conditions. When trained on clean data, its estimates for orbital decay u, velocity v, and precession δ closely match the ground truth. As noise is introduced, the model continues to generate smooth and realistic trajectories that effectively suppress most of the random scatter present in the training data. However, with increasing noise levels, a slight degradation in accuracy is observed. In particular, u tends to be underestimated, the amplitude of oscillations in v decreases, and cumulative phase shifts become more prominent in δ, especially at higher values of ϕ. Despite these deviations, the Neural ODE successfully captures the dominant physical trends and exhibits strong generalization ability in the presence of moderate noise. These results demonstrate that the model is robust to moderate noise but shows reduced precision in highly noisy scenarios, particularly in learning accumulated quantities like precession.



(a) no noise (b) moderate noise (c) high noise

Figure 2: UDE results for Case 1 across different noise levels.

From Figure 2, the UDE learns from data spanning the entire domain and performs remarkably well in capturing orbital dynamics across all noise levels. When the data is noise-free, the model's forecast aligns almost perfectly with the true trajectory throughout the full range of ϕ. Even when the training data exhibits moderate scatter due to noise, the UDE still produces clean and accurate predictions that closely follow the underlying orbital geometry. Under high noise, although the data becomes significantly more dispersed, the model's predictions remain smooth and physically plausible. Minor deviations appear, mostly in amplitude, but the overall trend of u is preserved. These results demonstrate that when given full access to the data, the UDE is highly effective at learning and retaining the correct orbital structure, even under substantial noise.



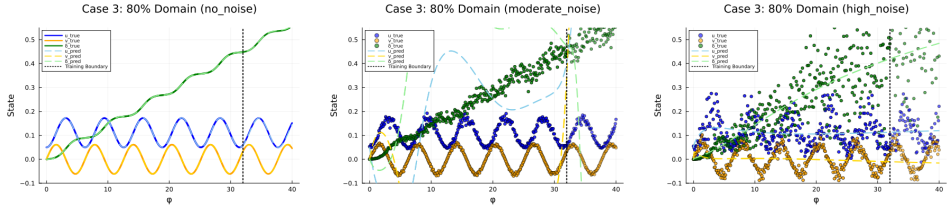
(a) no noise (b) moderate noise (c) high noise

Figure 3: Symbolic recovery results for Case 1 across different noise levels.

From Figure 3, symbolic recovery produces a nearly exact match between the learned correction and the true cubic relationship across all noise levels, as also confirmed by the metrics. With no noise, the learned expression is Correction$(u) \approx 0.009997\,u^3$ and the mean error is extremely low (6.1×10^{-9}), showing the method can recover the precise physical law. When moderate noise is added, the expression remains highly accurate at Correction$(u) \approx 0.009951\,u^3$, with a slight increase in the mean error (8.8×10^{-8}), but the symbolic curve still tracks the true correction almost perfectly. Even in the high-noise scenario, the recovered correction is Correction$(u) \approx 0.009964\,u^3$ and the mean error only rises to 6.5×10^{-8}. Across all conditions, the fit between learned and true corrections is visually tight throughout the full u range, and the metric values reflect the very strong noise robustness and fidelity of symbolic recovery when trained on the whole domain.

3.2 Case 3: Training on 80% of the orbital domain and forecasting

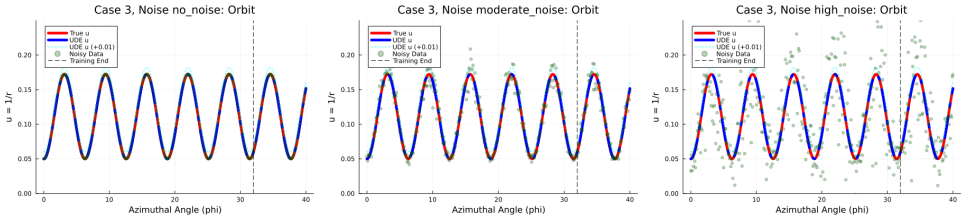
The Neural ODE and UDE models were also trained on 80% of the azimuthal domain and evaluated on the remaining 20% as the forecast region. The performance across different noise levels is illustrated in the following figures.



(a) no noise (b) moderate noise (c) high noise

Figure 4: Neural ODE results for Case 3 across different noise levels.

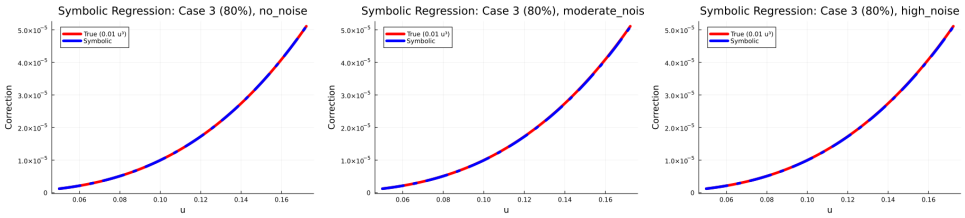
From Figure 4, where the Neural ODE is trained on 80% of the domain and the remaining 20% is utilized for prediction, the model is quite accurate when there is no noise. Its projected curves for orbital decay (u), velocity (v), and precession (δ) closely match the real solutions in both the training area and the forecast period, with almost total overlap between the prediction and reference. When moderate noise is present, the model continues to generate smooth and realistic predictions that follow the central trends and oscillatory behavior of the noisy data; only minor discrepancies and slight phase shifts arise, especially in the δ variable as ϕ increases into the forecast region. With high noise, the training data becomes much more dispersed, and the Neural ODE's outputs, while remaining physically plausible and smooth, increasingly diverge from the true solution: u is generally underestimated in the forecast domain, v's oscillations are reduced in amplitude, and δ displays noticeable phase errors, particularly beyond the boundary of the training set. Across all noise levels, the Neural ODE retains the main features and trends of the system, showing robustness to noise in its overall structure, but displaying greater quantitative error and accumulated deviations, especially for δ, when forecasting with highly corrupted data.



(a) no noise (b) moderate noise (c) high noise

Figure 5: UDE results for Case 3 across different noise levels.

From Figure 5, the UDE sees 80% of the domain and predicts the rest. It handles all noise levels well. When the data is clean, its prediction for orbital decay stays very close to the actual curve, both in the part it was trained on and the short region it has to forecast. The model generalizes well and captures the expected behavior even beyond the training boundary. The blue dashed UDE curve almost entirely overlaps the red genuine solution. For moderate noise, even though the noisy training points show some scatter, the UDE's prediction stays smooth and follows the main trend very well, keeping the overall structure and rise of u throughout the domain; only small changes show up, mostly as slight changes in amplitude in the forecast region. When there is a lot of noise, the training data spreads out much more, yet the UDE still gives a clean, physically plausible forecast that stays close to the real orbit. The forecast interval shows a slight underestimation of the peak values, but the basic monotonic growth and overall trend are still there. Overall, the UDE handles noise well and predicts beyond its training zone. Even when the data is messy, it sticks to the true motion and smooths out random scatter.



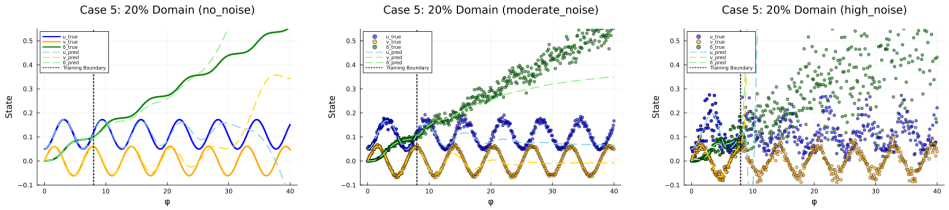
(a) no noise (b) moderate noise (c) high noise

Figure 6: Symbolic recovery results for Case 3 across different noise levels.

From Figure 6, for all noise levels, the symbolic recovery produces a learned correction curve that stays remarkably close to the true cubic correction throughout the range of u. When there is no noise, the learned expression matches the exact physical law (Correction$(u) \approx 0.010000\,u^3$) with an extremely small mean error, making the blue and red curves visually overlap across the plot. For moderate noise, the symbolic fit remains very accurate (Correction$(u) \approx 0.009961\,u^3$) and the mean error is still very low, so the learned and true curves are nearly indistinguishable. Even in the high-noise case, the learned coefficient is 0.009971 while the mean error is modest (5.2×10^{-8}), and the main trend and amplitude of the correction are preserved in the plot. These results highlight that, even as noise increases and only 80% of the domain is used for training, the symbolic recovery robustly extracts the intended cubic relationship, with minimal deviation from the true physical form.

3.3 Case 5: Training on 20% of the orbital domain and forecasting

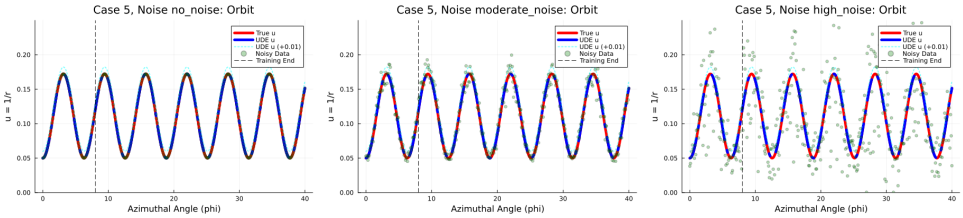
To test the model's extreme extrapolation capability, the Neural ODE and UDE models were trained on only 20% of the azimuthal domain. The evaluation results across different noise levels are presented in the following figures.



(a) no noise (b) moderate noise (c) high noise

Figure 7: Neural ODE results for Case 5 across different noise levels.

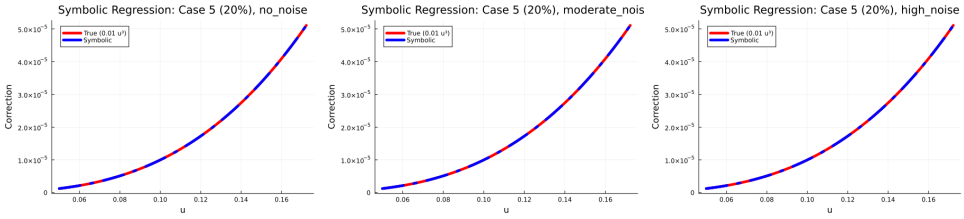
From Figure 7, when the Neural ODE is trained on only 20% of the domain, the model's ability to predict the other 80% depends significantly on the level of noise. In the no-noise case, the predictions for orbital decay u, velocity v, and precession δ roughly match the real paths in the training area. However, the forecasts for $\phi \approx 8$ show clear differences, especially with v and δ growing in amplitude and going out of phase. With moderate noise, the training data becomes scattered, yet the model maintains smooth trajectories that capture overall trends within the limited training region; however, forecasting errors increase substantially, showing flattening oscillations in v and underestimation of u, while δ deviates more markedly in phase during extrapolation. Under high noise, the training data is highly dispersed; the model's forecasted trajectories lose coherence, with u substantially underestimated after the training boundary, v's oscillations diminished significantly, and δ's phase errors growing rapidly, resulting in poor long-term predictive accuracy. Overall, while the Neural ODE handles interpolation within the sparse noisy training data reasonably well, its forecasting performance diminishes noticeably in both amplitude and phase with reduced data fraction and increased noise, especially affecting the cumulative precession variable δ.



(a) no noise (b) moderate noise (c) high noise

Figure 8: UDE results for Case 5 across different noise levels.

From Figure 8, the UDE model is trained using just 20% of the domain and tasked with forecasting the remaining 80%. When the data is noiseless, the UDE's prediction for orbital decay u follows the true curve almost exactly, not only within the limited training region but also far into the extended forecast area. This close tracking demonstrates that the model generalizes the expected behavior well even when trained on a small subset. With moderate noise, the training data shows visible scatter, but the UDE continues to produce a smooth and reliable curve. Although small amplitude differences and minor underestimation appear in the forecast region, the overall rising trend and physical structure of u are well preserved. Under high noise, the data is widely dispersed, yet the UDE manages to predict a smooth, stable orbit. Though peak values are more damped beyond training, the model still captures the main decay trend, reflecting its resilience and extrapolation ability.



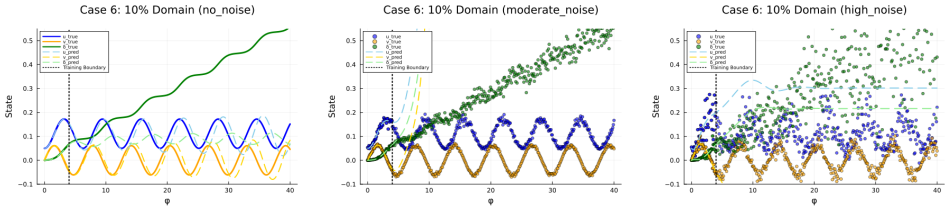
(a) no noise (b) moderate noise (c) high noise

Figure 9: Symbolic recovery results for Case 5 across different noise levels.

From Figure 9, where the symbolic correction is recovered using only 20% of the domain for training, the results remain exceptionally accurate under all noise conditions. The learned correction curve from the symbolic regression (blue line) is nearly indistinguishable from the true cubic correction (red line) across the full range of u, as seen in the images. When there is no noise, the recovered expression is almost exactly $\text{Correction}(u) \approx 0.009999\,u^3$, with a mean error of just 9.6×10^{-10}, indicating perfect agreement. With moderate noise, the learned correction ($0.009999\,u^3$) and mean error (1.8×10^{-9}) remain extremely close to the theoretical value, with the two curves overlapping entirely. Even with strong noise, the recovered term still comes out as $0.010000\,u^3$, and the mean error drops to 3.2×10^{-10}. The fit looks almost identical to the true correction curve. This shows that the method can still pull out the right cubic relation, staying very close to the actual physics, even when the data is noisy and limited.

3.4 Case 6: Training on 10% of the orbital domain and forecasting

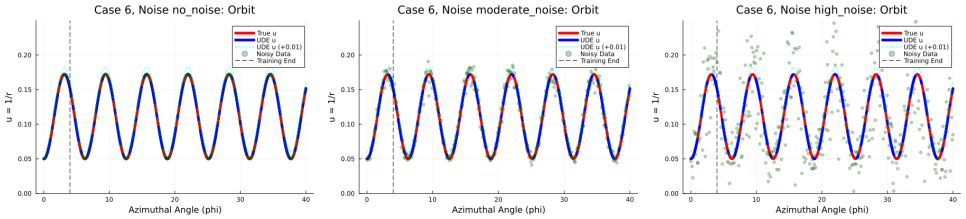
In this final case, the Neural ODE and UDE models were trained on only 10% of the azimuthal domain, representing the most limited training setting. Forecasting results under different noise conditions are shown in the following figures.



(a) no noise (b) moderate noise (c) high noise

Figure 10: Neural ODE results for Case 6 across different noise levels.

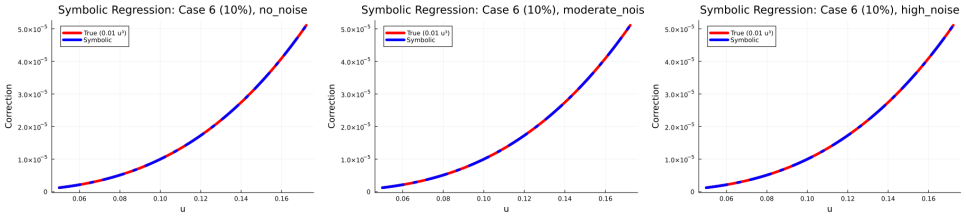
From Figure 10, where the Neural ODE is trained on just 10% of the domain and must forecast the remaining 90%, the model's predictive performance is strongly contingent on the noise present in the data. With no noise, the model's predicted curves for orbital decay (u), velocity (v), and precession (δ) follow the true solution closely within the short training region, but discrepancies emerge soon after, especially in the forecast interval: v and δ begin to drift in amplitude and phase, and u diverges more rapidly as ϕ increases. When moderate noise is present, the Neural ODE still fits the sparse and noisy training data well, producing smooth trajectories that follow the dominant trends in u and the oscillatory character in v and δ. However, as soon as the model forecasts beyond the small training window, errors increase sharply: u is underestimated, oscillations in v flatten quickly, and δ accumulates noticeable phase error, deviating from the true precession curve. Under high noise, the prediction becomes even less accurate; the model fits the noisy sample in the tiny training region but fails to generalize, u tracks too low, v's oscillations dampen to near-flatness, and δ departs rapidly from the true cumulative trend, demonstrating significant phase error. Even with limited training data, the Neural ODE captures basic dynamics well. But as noise increases and the forecast region grows, its accuracy drops, especially for δ, where small errors build up over time.



(a) no noise (b) moderate noise (c) high noise

Figure 11: UDE results for Case 6 across different noise levels.

From Figure 11, the UDE is trained with just 10% of the domain and must forecast the remaining 90%, making it the most challenging scenario. When there is no noise in the data, the UDE's prediction for orbital decay (u) nearly overlays the exact solution within the tiny training region and manages to follow the expected upward trend well into the much larger forecast region, with only minor discrepancies emerging as ϕ increases. With moderate noise, the training data are visibly scattered, but the UDE continues to produce a smooth and reasonable forecast; although slight underestimation and small amplitude differences are evident past the training boundary, the main trend is maintained. In the high-noise case, the scattered data points make the task especially difficult, yet the UDE still yields a clean and physically plausible prediction. There is some notable underestimation in the forecast region and further flattening of the rise, but the general monotonic behavior of u and the overall shape are preserved. These results show that, even when trained on a minimal portion of the domain and exposed to heavy noise, the UDE can still generalize and filter out noise, maintaining the essential trends of orbital decay, though forecasting accuracy naturally drops as data coverage and quality decrease.



(a) no noise (b) moderate noise (c) high noise

Figure 12: Symbolic recovery results for Case 6 across different noise levels.

From Figure 12, the symbolic recovery is still quite accurate even though it only has 10% of the domain for training and has to deal with all three noise circumstances. In all of the figures, the learned correction curve (blue) lines up almost exactly with the genuine cubic correction (red) in the no-noise, moderate-noise, and high-noise instances. The metrics back up this visual conclusion, with mean errors of only 1.9×10^{-9} (no noise), 4.0×10^{-10} (moderate noise), and 2.6×10^{-10} (high noise), all very tiny. Even when there is a lot of noise, the symbolic and true curves are not clearly separated. This shows that the symbolic regression can reliably find the right physical relationship ($\text{Correction}(u) = 0.01\,u^3$) no matter how little data there is or how much noise there is. This demonstrates the strength and reliability of the method, even in the hardest situation.

4 Discussion & Conclusion

This work benchmarks Neural ODEs and Universal Differential Equations (UDEs) for forecasting and symbolic recovery in the relativistic Schwarzschild orbital system under varying data and noise regimes. Neural ODEs, while effective at fitting training data and interpolating trajectories with low mean absolute errors (e.g., $\text{MAE}\langle u, v, \delta\rangle \approx 0.0004$ on the full, noiseless domain), require extensive coverage ($\geq$80%) for reliable forecasting; performance degrades sharply in data-limited regimes (e.g., forecast MAE for u rises above 0.026 when trained on 10% of the domain), especially as noise increases.

UDEs, in contrast, maintain low forecasting errors for both system state and correction term (`diff_corr` and `diff_orbit` often $\lesssim 10^{-7}$) even when trained with as little as 10–20% domain data, and show strong robustness to moderate and even high noise.

However, several limitations persist. Both approaches show reduced accuracy and stability when forecasting over long horizons, and performance can decline rapidly in the presence of extreme noise or highly non-representative training sets. Neural ODEs in particular may extrapolate unphysically outside their training domain [6, 7, 8], while UDEs, despite their robustness, still rely on the quality of the embedded physical model and the expressivity of the neural component to capture all relevant corrections [2]. Furthermore, the assessment here is limited to the Schwarzschild system; generalization to more complex or less structured problems remains an open question.

Crucially, across all experiments, symbolic regression recovered the physical correction term with very high fidelity (mean error $\lesssim 10^{-7}$). This confirms that UDEs can provide reliable, interpretable corrections while maintaining accuracy. Compared to Neural ODEs [6, 7, 8], UDEs [2] and physics-informed methods [4, 3] provide stronger generalization in noisy, data-scarce regimes, while preserving interpretability via symbolic discovery [13, 14, 15] and aligning with broader physics-informed ML trends [20].

Looking ahead, we aim to strengthen long-term stability and apply this approach to astrophysical problems where symbolic recovery could expose new relativistic or beyond-GR effects.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

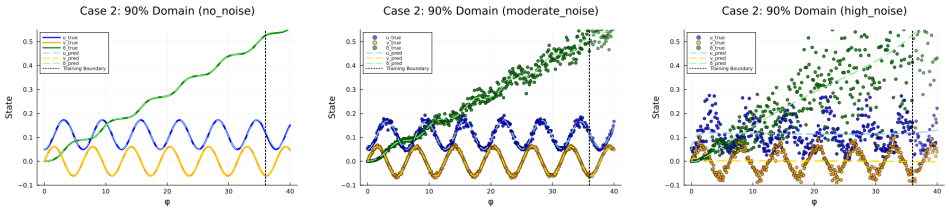
The authors thank colleagues at Delhi Technological University and Vizuara AI Labs for helpful discussions and feedback.

Appendix A: Additional results

This appendix presents the two additional cases (90% and 40% training coverage) omitted from the main text for brevity. Both follow the same experimental setup and evaluation protocol described in Section 2.

Case 2: Training on 90% of the orbital domain and forecasting

The Neural ODE and UDE models were trained on 90% of the azimuthal domain and evaluated on the remaining 10%. The results are shown in the following figures.

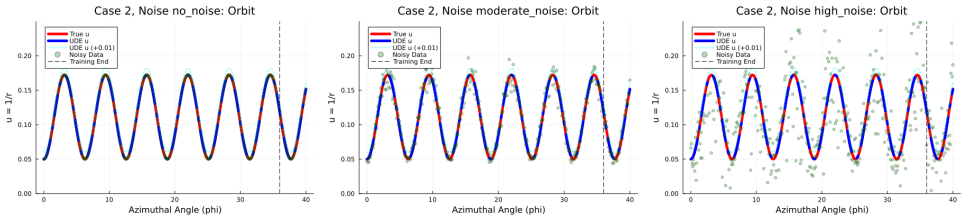


(a) no noise (b) moderate noise (c) high noise

Figure 13: Neural ODE results for Case 2 (90% training) across different noise levels.

From Figure 13, the Neural ODE performs well when no noise is present. Its predictions for orbital decay (u), velocity (v), and precession (δ) nearly overlap with the true curves, both in the training portion and in the last 10% used for forecasting. When moderate noise is added, the model still gives smooth and physically reasonable results, though in δ some small deviations and phase shifts

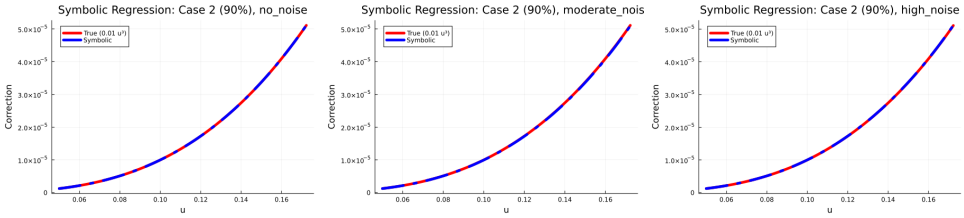
begin to show at larger values of ϕ. Under high noise, the predictions remain plausible and smooth but show noticeable underestimation of u, reduced oscillation amplitude in v, and more pronounced phase error in δ. Across all noise levels, the Neural ODE preserves the principal physical behavior and shows robustness to moderate noise, though errors accumulate in the forecast region at higher noise levels.



(a) no noise (b) moderate noise (c) high noise

Figure 14: UDE results for Case 2 (90% training) across different noise levels.

From Figure 14, the UDE continues to perform well across all noise levels. With no noise, its predictions for u overlap almost exactly with the true curve in both training and forecast regions. With moderate noise, the predictions remain smooth and closely track the underlying orbital behavior, effectively filtering out most scatter. Under high noise, the forecast slightly underestimates u beyond the training boundary, but the overall orbital structure and monotonic rise are preserved. These results confirm that the UDE generalizes reliably even when only 10% of the domain remains unseen.



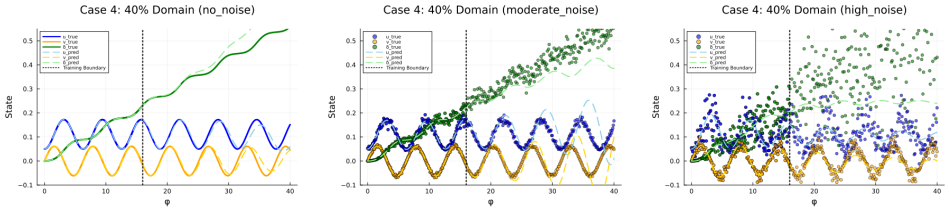
(a) no noise (b) moderate noise (c) high noise

Figure 15: Symbolic recovery results for Case 2 (90% training) across different noise levels.

From Figure 15, across all three noise levels, the symbolic recovery curve closely matches the true cubic correction. In the noiseless case, the recovered expression is $\text{Correction}(u) \approx 0.010000\,u^3$ with mean error 3.3×10^{-10}. With moderate noise, the expression remains highly accurate at $0.009964\,u^3$ with mean error 6.4×10^{-8}. Even with high noise, the recovery yields $0.010010\,u^3$ with mean error 1.9×10^{-8}. These results demonstrate that symbolic regression faithfully reconstructs the underlying correction term with strong robustness and accuracy, even when noise is high and only 90% of the domain is available for training.

Case 4: Training on 40% of the Orbital Domain and Forecasting

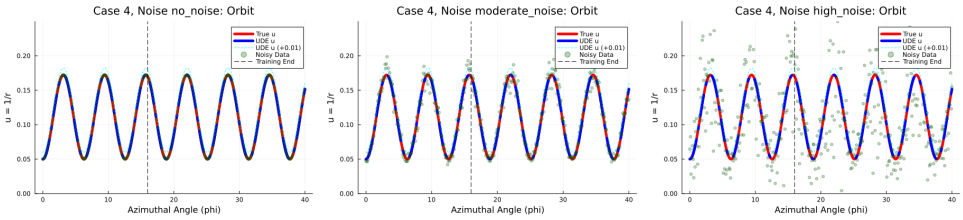
To evaluate long-range forecasting capability, the Neural ODE and UDE models were trained on only 40% of the azimuthal domain. The performance under varying noise conditions is shown in the following figures.



(a) no noise (b) moderate noise (c) high noise

Figure 16: Neural ODE results for Case 4 (40% training domain) across different noise levels.

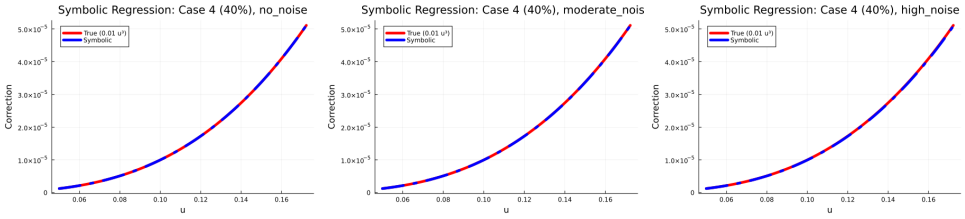
From Figure 16, the Neural ODE is trained on just 40% of the data and needs to predict the rest. Its accuracy depends heavily on how noisy the training data is. When there is no noise, the model's predictions for u, v, and δ are almost identical to the true curves, both within and beyond the training region. Even with moderate noise, it still performs well, capturing the upward trend in u and the wave-like behavior in v and δ, even far into the forecast area. There are some small differences though, like phase shifts in δ as ϕ increases, but overall, the model filters out the noise effectively and sticks to the correct trend. Under high noise, the spread in the training data becomes pronounced, and the Neural ODE's predictions, while still smooth and physically plausible, show larger discrepancies: u is generally underestimated in the forecast region, v's amplitude and regularity decrease, and δ accumulates increasingly significant phase errors as ϕ grows. Across all noise conditions, the Neural ODE preserves the primary trends and physical behaviors, but its forecast accuracy diminishes with increasing noise and data scarcity, especially for δ, which remains the most sensitive to error accumulation.



(a) no noise (b) moderate noise (c) high noise

Figure 17: UDE results for Case 4 (40% training domain) across different noise levels.

From Figure 17, where the UDE is trained on just 40% of the domain and forecasts the remaining 60%, the model demonstrates strong forecasting ability and noise robustness. When there is no noise in the data, the UDE's prediction for orbital decay (u) matches the true solution nearly perfectly, both within the training region and throughout the long forecast interval, with the predicted and actual curves closely following each other across the full range of ϕ. With moderate noise, the model continues to perform well: even as the training data shows noticeable scatter, the UDE's predicted curve remains smooth and closely tracks the underlying trend. Some small differences arise, mainly as subtle amplitude shifts or slight underestimation in the forecast region, but the main rising pattern of u is preserved. Under high noise, the data points are much more dispersed, yet the UDE still outputs a stable and physically reasonable curve. In the forecast zone, the model slightly smooths out the oscillations and tends to underestimate values, but it still captures the steady growth and the correct orbital pattern. Case 4 shows that even with limited training data and high noise, the UDE can pick up the key features of decay and stay reliable farther out.



(a) no noise (b) moderate noise (c) high noise

Figure 18: Symbolic recovery results for Case 4 across different noise levels.

From Figure 18, where only 40% of the domain is used for training to get the symbolic correction back, the results are still very accurate at all noise levels. In the situation with no noise, the learned correction curve (blue) is almost the same as the genuine cubic correction (red). This is shown by the mean error of only 3.5×10^{-9}, and the two curves overlap across the whole range of u. The symbolic curve still closely follows the genuine correction, with a mean error of 2.7×10^{-8}, even when there is substantial noise. The plot shows no apparent deviation; the predicted curve remains nearly on top of the reference, with a mean error of 7.7×10^{-8}, even under noisy conditions. The red and blue lines are virtually indistinguishable. These results confirm that symbolic regression reliably extracts the underlying cubic structure, preserving the correct physics despite less and noisier data.

References

[1] Christopher Rackauckas and Qing Nie. Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia. *Journal of Open Research Software*, 5(1):15, 2017.

[2] Christopher Rackauckas, Yingbo Ma, Jan Martensen, Collin Warner, Konstantin Zubov, Rohit Supekar, Dominic Skinner, and Ali Ramadhan. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

[3] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.

[4] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[5] Steven L Brunton and J Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2020.

[6] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[7] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. In *Advances in Neural Information Processing Systems*, 2019.

[8] Stefano Massaroli, Michael Poli, Giulio Bin, Atsushi Yamashita, and Hajime Asama. Dissecting neural odes. In *Advances in Neural Information Processing Systems*, 2020.

[9] Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam M Oberman. How to train your neural ode. *arXiv preprint arXiv:2002.02798*, 2020.

[10] Hanshu Yan, Jiawei Du, Vincent YF Tan, and Jiashi Feng. On robustness of neural ordinary differential equations. *arXiv preprint arXiv:1910.05513*, 2019.

[11] Charles W Misner, Kip S Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, 1973.

[12] Bernard F Schutz. *A First Course in General Relativity*. Cambridge University Press, 2 edition, 2009.

[13] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[14] Silviu-Marian Udrescu and Max Tegmark. Ai feynman: a physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

[15] Miles Cranmer, Rui Xu, Peter Battaglia, Shirley Ho, David Spergel, et al. Discovering symbolic models from deep learning with inductive biases. In *Advances in Neural Information Processing Systems*, 2020.

[16] Chris Rackauckas, Mike Innes, Yingbo Ma, Jesse Bettencourt, Lyndon White, and Vaibhav Dixit. Diffeqflux.jl: A julia library for neural differential equations. *arXiv preprint arXiv:1902.02376*, 2019.

[17] Mike Innes, Alan Edelman, Keno Fischer, Chris Rackauckas, Elliot Saba, Viral B Shah, and Will Tebbutt. A differentiable programming system to bridge machine learning and scientific computing. *arXiv preprint arXiv:1907.07587*, 2019.

[18] Christopher Rackauckas, Yingbo Ma, and collaborators. Diffeqflux.jl: Neural differential equations in julia. https://github.com/SciML/DiffEqFlux.jl, 2025. Accessed 2025-08-21.

[19] Christopher Rackauckas and the SciML Community. Sciml: Scientific machine learning open source software. https://sciml.ai, 2025. Accessed 2025-08-21.

[20] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.