

Towards Foundation Inference Models that Learn ODEs In-Context

Maximilian Mauel², Manuel Hinz¹², Patrick Seifner¹²,
David Berghaus¹³ & Ramsés J. Sánchez¹²³

¹Lamarr Institute, ²University of Bonn, ³Fraunhofer IAIS

seifner@cs.uni-bonn.de

Abstract

Ordinary differential equations (ODEs) describe dynamical systems evolving deterministically in continuous time. Accurate data-driven modeling of systems as ODEs, a central problem across the natural sciences, remains challenging, especially if the data is sparse or noisy. We introduce **FIM-ODE** (Foundation Inference Model for ODEs), a pretrained neural model designed to estimate ODEs *zero-shot* (i.e. *in-context*) from sparse and noisy observations. Trained on synthetic data, the model utilizes a flexible neural operator for robust ODE inference, even from corrupted data. We empirically verify that **FIM-ODE** provides accurate estimates, on par with a neural state-of-the-art method, and qualitatively compare the structure of their estimated vector fields.

Keywords: System Identification, ODE Inference, Foundation Models, In-Context Learning, Neural Operators, Simulation-based Inference

1 Introduction

Ordinary differential equations (ODEs) are a class of continuous-time dynamical systems, modeling phenomena evolving deterministically in continuous space. They are formally described by *vector fields* $\mathbf{f} : \mathbb{R}_+ \times \mathbb{R}^D \rightarrow \mathbb{R}^D$, defining the differential equation

$$d\mathbf{x}(t) = \mathbf{f}(t, \mathbf{x}(t)) dt, \quad (1)$$

with continuous solution trajectories $\mathbf{x} : \mathbb{R}_+ \rightarrow \mathbb{R}^D$. The ubiquitous presence of these simple models throughout the natural sciences is remarkable, from describing Newton’s laws of motion, to population dynamics in biological systems [Lotka, 1925, Volterra, 1926] and atmospheric convection in meteorology [Lorenz, 1963]. Accurate ODE models characterize the underlying phenomena (e.g. by fixed points or limit cycles) and can forecast future states.

In this work, we consider the *ODE system identification problem*: estimating the ODE (i.e. a vector field) that best describes a system from time series observations only. Traditional approaches to this problem are either *non-parametric* [Heinonen et al., 2018] or *symbolic-regression-based* [Dong and Zhong, 2025]. Recently, d’Ascoli et al. [2024] introduced **ODEFormer**, a pretrained *neural symbolic regression* method, transforming (tokenized) time series observations into (tokenized) equations.

Foundation Inference Models (FIMs) are a general simulation-based framework for inference of dynamical systems. These deep neural network models are pretrained on synthetic data, generated from a class of dynamical systems, using a *supervised* train objective to *match* the target and estimated infinitesimal generators of the processes. This framework has shown promising results in the inference of continuous-time Markov chains [Berghaus et al., 2024], stochastic differential equations [Seifner et al., 2025a], and point processes [Berghaus et al., 2025]. If appropriate, it defaults to *neural operators* [Kovachki et al., 2023] for continuous function estimates [Seifner et al., 2025b].

In this work, we introduce FIM-ODE, a FIM for the ODE identification problem, which uses a *neural operator* to model the vector field. Our model performs better than ODEFormer on a synthetic test set and we argue that FIM-ODE’s flexible neural operator yields more reasonable global vector field predictions.

2 Foundation Inference Model

The ODE inference problem is a special case of the stochastic differential equation (SDE) inference problem, tackled by FIM-SDE [Seifner et al., 2025a]. However, non-stochastic trajectories *do not explore the space as much as stochastic trajectories*, providing less information about the global vector field. Nevertheless, because of the problem similarities, we stick to the architecture of FIM-SDE.

Synthetic Data Generation. Following Seifner et al. [2025a], we generate synthetic training data from a broad distribution over ODE systems and initial states of up to dimension $D = 3$. Each component of a vector field is sampled from a distribution over multivariate polynomials of up to degree 3. We simulate these systems, record the solutions at (irregular) grids and corrupt them by additive noise to generate time series observations from them.

Inference Model Architecture. FIM-ODE estimates a vector field $\hat{\mathbf{f}}$ from a set $\mathcal{D} = \{\mathbf{y}_k\}_{k=1}^K$ of time series $\mathbf{y}_k = [(t_{k1}, \mathbf{y}_{k1}), \dots, (t_{kL}, \mathbf{y}_{kL})]$ *in-context*, without any further training or finetuning. The vector field estimate $\hat{\mathbf{f}}$ is implemented by a *DeepONet* neural operator [Lu et al., 2021]. The *branch-net* of FIM-ODE, a Transformer encoder, encodes $\mathcal{D}$ into $K(L-1)$ E -dimensional representations $\mathbf{D} \in \mathbb{R}^{E \times K(L-1)}$. Note that this retains individual encodings for almost all observations, which a *combination network* (see below) can access directly. The *trunk-net* is a linear map, encoding a location $\mathbf{x} \in \mathbb{R}^D$ into $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^E$. The *combination network* is a sequence of residual attention layers, using $\mathbf{D}$ as keys and values, and $\mathbf{h}(\mathbf{x})$ as query, similar to a Transformer decoder. A final linear projection yields the estimate $\hat{\mathbf{f}}(\mathbf{x})$.

Training. Adhering to the FIM framework, we use a *supervised training objective* $\mathcal{L}(\mathbf{x}, \mathcal{D}, \mathbf{f}) = \|\hat{\mathbf{f}}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\|^2$, matching the predicted to the ground-truth vector fields on $\mathbf{x}$ sampled “close” to the observation values in $\mathcal{D}$. We pretrain a single FIM-ODE with roughly 20M parameters on 600k synthetic equations¹.

¹For comparison, ODEFormer is trained on 50M equations and has 86M parameters.

Table 1: Comparison of R^2 -accuracy on polynomial vector field data, given a single trajectory as context.

Model	Reconstruction Task	Generalization Task
ODEFormer	0.65	0.19
FIM-ODE	0.90	0.26

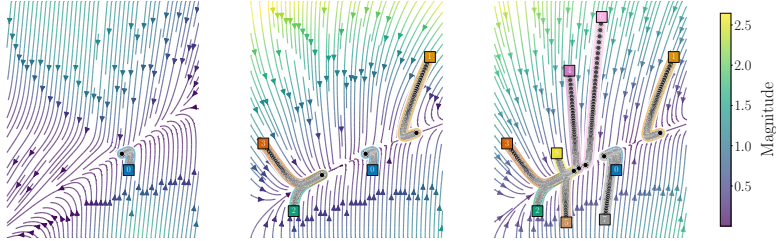


Figure 1: Vector field estimates of FIM-ODE from 1, 5 and 9 context trajectories of the same system. The estimates improve, given more trajectories.

3 Experiments

For our experiments, we sample 4000 ODEs with polynomial vector fields of up to degree 3 and up to dimension 3. We simulate 9 trajectories with initial states sampled from $\mathcal{N}(0, 1)$ from each ODE, recording 200 observations on a regular grid with inter-observation times $\Delta\tau = 0.05$. Following d’Ascoli et al. [2024], we measure performance by the R^2 -accuracy, the percentage of R^2 scores larger than 0.9. We compare our pretrained FIM-ODE to a pretrained ODEFormer² on two tasks: a *reconstruction task*, in which performance is measured on the reconstruction of the *context trajectory*, and a *generalization task*, in which performance is measured on the reconstruction of *held-out trajectories*. Note that ODEFormer can only process a single context trajectory, so we restrict to that case in direct comparisons.

Reconstruction and Generalization Performance. Table 1 contains the R^2 -accuracy for both tasks on the polynomial test set. Note that the polynomial test data is in the training distribution of both models, but ODEFormer was trained on a broader distribution than FIM-ODE. Our model performs well in both tasks, even better than ODEFormer. The large difference in R^2 -accuracy shows that generalization is inherently more difficult than reconstruction.

Multi-trajectory Context. FIM-ODE is designed to process multiple trajectories of the same system simultaneously. Figure 1 depicts vector field estimates of FIM-ODE with varying counts of context trajectories. For a single

²Source of pretrained ODEFormer weights: <https://github.com/sdascoli/odeformer>

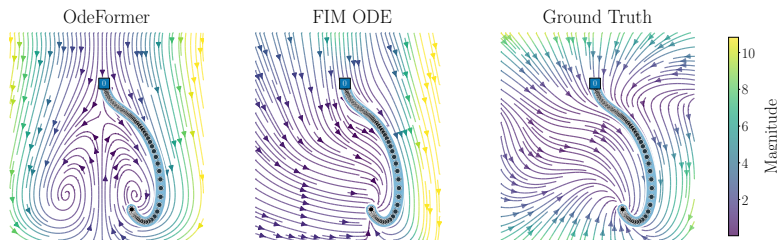


Figure 2: Comparison of vector field estimates to the ground-truth, given a single trajectory. ODEFormer predicts a complex global pattern, while FIM-ODE focuses on accurate predictions near the provided trajectory.

trajectory, vector field estimates are inaccurate at locations distant from the observations. With more trajectories covering the space, FIM-ODE corrects these estimates, extracting and combining all available information effectively. Such accurate global vector field prediction is particularly beneficial for generalization tasks — as well as for *interpretability in scientific settings*.

Local and Global Predictions. In Figure 2, we visually compare vector field estimates of FIM-ODE and ODEFormer from a single trajectory. Locally, at the observations, FIM-ODE predicts a complex pattern to reconstruct the trajectory. Globally, far away from the observations, the predictions are much simpler. Compared to the ground-truth, ODEFormer predicts a more complex vector field. This results in complex global pattern predictions, largely unjustified by the single, quite simple observed trajectory. The stark difference between the two models is explained by the different vector field parametrizations. ODEFormer is restricted to predict (rational) polynomial symbolic equations, which, given sparse or noisy observations, might not default to simple expressions. *The neural operator of FIM-ODE can handle these situations more flexible.*

4 Conclusions

In this work, we introduced FIM-ODE, a pretrained foundation inference model that estimates ODEs from time series observations, utilizing a neural operator. In our preliminary experiments, it demonstrates better performance than ODEFormer, a state-of-the-art neural symbolic regression approach. We find structural differences between the vector fields predicted by these approaches and argue that FIM-ODE’s global predictions are more justifiable.

In future work, we will compare these methods on ODEBench, a ODE inference benchmark dataset of 63 hand-selected ODEs. We will also explore the discovery of latent dynamics using a pretrained FIM-ODE, with possible applications in neural population and chemical reaction dynamics [Duncker et al., 2019], and the evolution of natural language content [Cvejosi et al., 2023, 2022].

Acknowledgments

This research has been funded by the Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia as part of the Lamarr Institute for Machine Learning and Artificial Intelligence.

References

- D. Berghaus, K. Cvejoski, P. Seifner, C. Ojeda, and R. J. Sánchez. Foundation inference models for markov jump processes. In *Advances in Neural Information Processing Systems*, volume 37, 2024.
- D. Berghaus, P. Seifner, K. Cvejoski, C. Ojeda, and R. J. Sánchez. In-context learning of temporal point processes with foundation inference models, 2025. URL <https://arxiv.org/abs/2509.24762>.
- K. Cvejoski, R. J. Sánchez, and C. Ojeda. The future is different: Large pre-trained language models fail in prediction tasks. *arXiv preprint arXiv:2211.00384*, 2022.
- K. Cvejoski, R. J. Sánchez, and C. Ojeda. Neural dynamic focused topic model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12719–12727, 2023.
- S. d’Ascoli, S. Becker, P. Schwaller, A. Mathis, and N. Kilbertus. ODEFormer: Symbolic regression of dynamical systems with transformers. In *The Twelfth International Conference on Learning Representations*, 2024.
- J. Dong and J. Zhong. Recent advances in symbolic regression. *ACM Comput. Surv.*, 57(11), 2025. ISSN 0360-0300.
- L. Duncker, G. Böhner, J. Boussard, and M. Sahani. Learning interpretable continuous-time models of latent stochastic dynamical systems. In *International conference on machine learning*, pages 1726–1734. PMLR, 2019.
- M. Heinonen, Çağatay Yıldız, H. Mannerström, J. Intosalmi, and H. Lähdesmäki. Learning unknown ode models with gaussian processes. In *International Conference on Machine Learning*, 2018.
- N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89): 1–97, 2023. URL <http://jmlr.org/papers/v24/21-1524.html>.
- E. N. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130 – 141, 1963.
- A. J. Lotka. *Elements of Physical Biology*. Williams & Wilkins Co., Baltimore, MD, 1925.

- L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.
- P. Seifner, K. Cvejoski, D. Berghaus, C. Ojeda, and R. J. Sanchez. Foundation inference models for stochastic differential equations: A transformer-based approach for zero-shot function estimation, 2025a. URL <https://arxiv.org/abs/2502.19049>.
- P. Seifner, K. Cvejoski, A. Körner, and R. J. Sanchez. Zero-shot imputation with foundation inference models for dynamical systems. In *The Thirteenth International Conference on Learning Representations*, 2025b.
- V. Volterra. Variazioni e fluttuazioni del numero d’individui in specie animali conviventi. *Memorie della Reale Accademia Nazionale dei Lincei*, 2:31–113, 1926.