Supervised Fine-Tuned Transformers for Symbolic Discovery of Power System Dynamics

Lu Sun and Nanpeng Yu
Department of Electrical and Computer Engineering
University of California, Riverside
Riverside, California 92507 USA
lsun098@ucr.edu, nyu@ece.ucr.edu

***Abstract*—Accurate dynamic models are essential for reliable power system stability analysis. However, practical models often deviate from real-world behavior due to simplifying assumptions and uncertain parameters. This paper proposes a supervised fine-tuned transformer for symbolic discovery of power system dynamics from post-fault transient trajectories by supervised fine-tuning ODEFormer, a transformer-based model for symbolic regression. A systematic data generation procedure is developed to simulate multi-machine systems under diverse operating conditions, fault locations, and durations. Numerical studies are conducted on 2-generator and 3-generator systems to evaluate the effectiveness of the proposed framework under different training and inference settings. The results demonstrate that the proposed method achieves high structural correctness and accurate trajectory reconstruction across different system configurations. In addition, mixed-topology training experiments indicate that when training data are insufficient, higher-dimensional system samples can improve the trajectory and parameter accuracy, while lower-dimensional system samples can improve the structure correctness and interpretability. These findings highlight the potential of transformer-based models for symbolic learning and interpretable modeling of power system dynamics.**

***Index Terms*—Power system dynamics, symbolic regression, transformer.**

I. Introduction

Running power system dynamic simulations is essential for system operators to assess stability and evaluate system response under a wide range of contingencies. The fidelity of these simulations critically depends on the accuracy of the underlying dynamic models. In practice, however, the models used in operational studies often deviate from real-world system behavior due to simplifying assumptions, incomplete representations of control systems, or inaccurate parameter settings. These discrepancies can reduce the robustness of stability assessments, particularly under stressed or abnormal operating conditions.

This gap highlights the need for data-driven approaches that can learn both the functional forms and parameters of power system dynamics directly from high-resolution measurement data. The challenge is especially pronounced during post-fault conditions, where network topology changes and hidden or proprietary control actions can significantly alter system dynamics in ways not captured by conventional models. In this paper, we propose a supervised fine-tuned transformer to extract interpretable differential equations from transient trajectories that better reflect real-world system behavior by supervised fine-tuning ODEFormer, a transformer-based model for symbolic regression.

There are two main research directions in data-driven power system dynamic analysis. The first focuses on leveraging machine learning models to directly predict dynamic trajectories from sensor measurements without relying on explicit physical models. Representative examples include Fourier neural operator–based frequency-domain predictors [1], machine learning–based reduced-order models [2], and neural ordinary differential equation (ODE)–based decentralized forecasters [3]. While these approaches demonstrate strong predictive performance, they often lack interpretability, do not recover explicit governing equations, and may violate physical constraints when operating outside the training distribution. In addition, their dependence on large volumes of training data limits their effectiveness in settings with sparse or partially observed measurements.

A second line of research focuses on dynamic parameter estimation using phasor measurement unit (PMU) data. Existing PMU-based parameter estimation methods include time-domain trajectory fitting, frequency-domain modal matching, and hybrid signal-processing-assisted estimation. For example, [4] estimates unknown generator parameters by fitting simulated trajectories to PMU measurements with a physics-based neural ODE, [5] identifies inertia and damping parameters by matching model eigenvalues and mode shapes to PMU-derived modal characteristics, and [6] combines ESPRIT-based oscillation extraction with weighted nonlinear least squares to estimate system inertia and related dynamic parameters from event-driven PMU data. [7] proposes a Nearly-Hamiltonian neural network to predict transient trajectory and dynamic parameters of power systems by embedding energy conservation laws in the proposed neural network architecture. These approaches assume that the underlying dynamic model structure is known a priori, and focus solely on estimating unknown parameters. Such an assumption limits their applicability when the true system dynamics deviate from simplified modeling assumptions or when control mechanisms are partially unknown.

To address these limitations, symbolic learning methods offer a promising approach for discovering governing equations directly from power system trajectory data [8]. For example, [9] extends SINDy-based symbolic regression for online identification of synchronous generator (SG) dynamics, improving the handling of noisy measurements, nonlinearities, and physical constraints. However, this approach remains a gray-box method, as it relies on a hand-crafted candidate library and a pre-specified SG model structure. Moreover, it is limited to local measurements and does not examine whether the learned symbolic dynamics generalize across different network topologies.

A transformer-based foundation model, ODEFormer, demonstrated the ability to recover multidimensional ODEs from noisy and irregularly sampled time series, highlighting the potential of transformer architectures for interpretable equation discovery from trajectory data [10], [11]. However, ODEFormer faces several challenges. For example, different symbolic expressions may have the same output on all possible initial states, which means that it might be difficult to get the ODEs with the ideal power system dynamic structure. It also struggles with high-dimensional ODE systems, as sequence length increases linearly with the system dimension, which may impair scalability for larger-dimensional systems. In addition, the training ODEs used in ODEFormer are synthetically generated from random symbolic expressions and do not explicitly represent networked systems with sinusoidal functions. These limitations highlight the need for a power system–specific symbolic learning framework that adapts transformer-based equation discovery to networked swing dynamics and disturbance-driven trajectories.

The limitations of existing work motivate the problem studied in this paper: symbolic discovery of power system dynamics from post-fault transient trajectories using sensor measurements. Unlike prior approaches that either predict trajectories without explicit equations or estimate parameters within fixed model structures, we aim to recover interpretable governing equations directly from data. Furthermore, our focus extends beyond single-machine models to multi-machine power system dynamics.

The main contributions of this work are as follows:

- We develop a transformer-based model for symbolic discovery of power system dynamics from post-fault transient trajectories by supervised fine-tuning ODEFormer on power system-specific datasets.
- A systematic data generation procedure is developed to simulate multi-machine systems under diverse operating conditions, fault locations, and durations.
- The proposed supervised fine-tuned transformer achieves high structural correctness and reliable trajectory reconstruction across different system configurations.
- We demonstrate that, under limited training data, cross-topology training can improve symbolic discovery performance: higher-dimensional system samples enhance trajectory and parameter accuracy, while lower-dimensional system samples improve structural correctness and interpretability.

The remainder of the paper is organized as follows. Section II presents the problem formulation. Section III describes the proposed model developed by supervised fine-tuning ODEFormer on power system-specific datasets. Section IV provides the numerical study setup and results. Finally, Section V concludes the paper.

II. Problem Formulation

This work aims to develop a supervised fine-tuned transformer to infer symbolic expressions of power system dynamics from disturbance-driven time-domain data. This approach provides a data-driven pathway to bridge high-fidelity simulations or real-world sensor data and compact analytical models, enabling improved interpretability, generalization, and potential insights into the underlying physical mechanisms. In this section, we will describe the problem setup, followed by an introduction to the classical models of power system dynamics.

A. Problem Setup

Assuming that the time-series trajectories of generator phase angles are available from simulations or measurements, the phase angle of generator i at time t is given by $\delta_i(t)$, and the rotor speeds $\omega_i(t)$ are sequentially obtained from the phase angle trajectories through numerical differentiation. Here, the subscript $i = 1, 2, \ldots, n$ indexes the generators, and $t = 1, 2, \ldots, T$ indexes the time steps.

By concatenating the phase angles and rotor speed deviations, the sampled system state at time t is defined as (1).

$$x(t) = [\delta_1(t), \omega_1(t), \delta_2(t), \omega_2(t), \ldots, \delta_n(t), \omega_n(t)]^T \quad t = 1, 2, \ldots, T \tag{1}$$

The state trajectory of an n-generator power system is then given by (2):

$$D = \{x(t)\}_{t=1}^{T}. \tag{2}$$

Given the trajectories D, the goal is to learn the underlying dynamic equations governing the system, which can be written in compact form as (3):

$$\dot{x} = f(x), \tag{3}$$

where $x = [\delta_1, \omega_1, \delta_2, \omega_2, \ldots, \delta_n, \omega_n]^T$.

B. Classical Swing Equations and Dynamic Simulation

In this work, we consider a n-generator power system described by the classical power system dynamic model. This model offers a physically interpretable and computationally efficient framework that captures essential nonlinear electromechanical dynamics while maintaining a manageable system dimension.

The classical multi-machine model uses a set of differential equations to describe the dynamic of the generators, where each equation is essentially the swing equation of generator i $(i = 1, 2, \ldots, n)$ as shown in (4):

$$M_i \frac{d^2\delta_i}{dt^2} + D_i \frac{d\delta_i}{dt} + P_{Gi}(\delta_i) = P_{Mi}^0 \quad i = 1, 2, \ldots, n, \tag{4}$$

where M_i and D_i respectively denote the inertia constant and the damping coefficient of generator i, $P_{Gi}(\delta_i)$ is the electric power output at phase angle δ_i, and P_{Mi}^0 is the mechanical power input at initial phase angle δ_i^0.

The electric power output of each generator is given by the power flow equations in terms of the admittance matrix parameters from Kron reduced Y-bus matrix ($Y_{ik} = G_{ik} + jB_{ik}$), which is shown in (5):

$$P_{Gi} = \sum_{k=1}^{n} |E_i||E_k|(G_{ik}\cos(\delta_i - \delta_k) + B_{ik}\sin(\delta_i - \delta_k)) \qquad i = 1, 2, \ldots, n. \tag{5}$$

Substituting (5) for P_{Gi} in (4), neglecting the damping coefficient, and converting equations into a set of coupled first-order differential equations, we have (6) for a multi-machine system.

$$\begin{aligned} M_i\dot{\omega}_i &= P_{Mi}^0 - |E_i|^2 G_{ii} \\ &\quad - \sum_{\substack{k=1 \\ k\neq i}}^{n} |E_i||E_k|[B_{ik}sin(\delta_i - \delta_k) + G_{ik}cos(\delta_i - \delta_k)] \\ \dot{\delta}_i &= \omega_i \qquad\qquad i = 1, 2, \ldots, n \end{aligned} \tag{6}$$

Since the fault condition essentially affects the system topology and parameters, the Kron-reduced admittance matrix Y needs to be calculated respectively based on pre-fault, faulted and post-fault conditions, denoted as Y_{pre}, $Y_{faulted}$, and Y_{post}. To simulate the system dynamics over the considered time interval, the corresponding ODEs are accordingly constructed and integrated sequentially, with the feasible pre-fault power flow solution chosen as the initial state. Only the post-fault trajectories are used for model training, while the pre-fault and faulted stages are simulated only to generate physically consistent initial conditions for the post-fault dynamics.

III. Technical Methods

A. Overall Framework

The overall workflow of the supervised fine-tuned transformer for symbolic discovery of power system dynamics is illustrated in Fig. 1, which consists of a supervised fine-tuning phase and an inference phase. In the fine-tuning phase, multi-machine power system trajectories are generated through power flow analysis and power system dynamic simulations under fault conditions, and the resulting trajectories and symbolic dynamic equations are tokenized as supervised input-output pairs. Numerical constants are encoded using sign, mantissa, and exponent tokens. A pretrained ODEFormer is then fine-tuned with cross-entropy loss, where the encoder processes trajectory inputs and the decoder generates symbolic equation tokens. In the inference phase, the fine-tuned model takes unseen trajectory data as input and outputs candidate symbolic equations. Greedy decoding is applied first, followed by beam sampling if the decoded expression is not interpretable.

B. Supervised Fine-tuning of ODEFormer

ODEFormer is a transformer-based symbolic regression model designed to infer the governing ODEs of dynamical systems from observed trajectories. Given time-series observations of $x(t)$, the model aims to recover the symbolic expression of the underlying dynamics $\dot{x}(t) = f(x(t))$. ODEFormer formulates this task as a sequence generation problem and adopts an encoder–decoder transformer architecture in which the encoder processes the observed trajectory, while the decoder automatically generates the symbolic representation of the ODE. The attention mechanism is the core operation inside these transformer blocks. Specifically, the scaled dot-product attention is defined as (7):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{7}$$

where Q, K, and V respectively denote the query, key, and value matrices, and d_k is the dimension of the key vectors.

We fine-tuned the pretrained ODEFormer model using simulation data generated from the classical power system dynamic model. These trajectory-ODE pairs serve as supervision for adapting the pretrained model to better capture the dynamics of networked systems characterized by sinusoidal coupling functions. Numerical constants are encoded using scientific notation, where each number is decomposed into tokens representing its sign, mantissa, and exponent, allowing continuous values to be generated through discrete token sequences.

During fine-tuning, the transformer model parameters are updated by minimizing the cross-entropy loss between the predicted token distribution and the ground-truth symbolic expression, as formulated in (8):

$$Loss = -\frac{1}{Z}\sum_{b=1}^{B}\sum_{s=1}^{S} w_{bs}\sum_{c=1}^{C} y_{bs,c}\log(\hat{y}_{bs,c}), \tag{8}$$

where B, S, and C denote batch size, maximum sequence length, and vocabulary size, respectively. $\hat{y}_{bs,c}$ represents the predicted probability of the s-th token of the b-th batch being c, and $y_{bs,c}$ is the one-hot ground truth. The term w_{bs} is a binary mask used for padding, and Z is the normalization constant.

During inference, a greedy decoding strategy is first applied to generate a candidate symbolic expression. The decoded expression is then checked for interpretability. If the generated expression is not interpretable, beam sampling with a beam size of 5 is subsequently employed to explore alternative candidates. In beam sampling, the decoding randomness is controlled by a temperature parameter τ, which rescales the logits before the softmax operation as shown in (9):

$$p_i = \frac{\exp(z_i/\tau)}{\sum_{j=1}^{C}\exp(z_j/\tau)}, \tag{9}$$

where z_i is the logit of the i-th token, C is the vocabulary size, and p_i is the corresponding sampling probability. A smaller τ sharpens the probability distribution and makes decoding closer to greedy search, whereas a larger τ flattens the distribution and increases exploration.

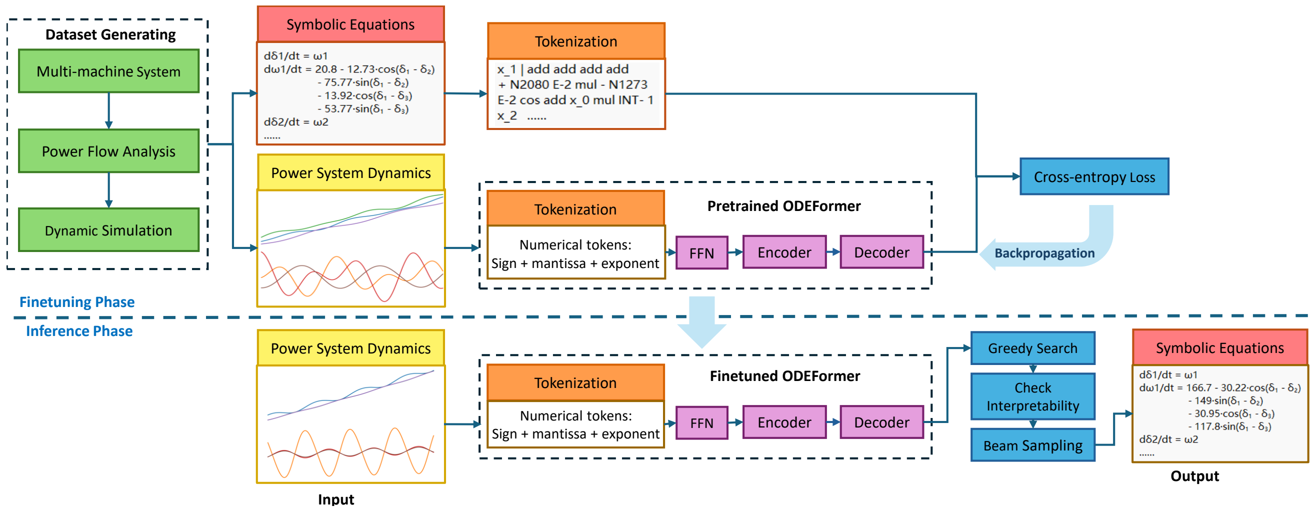


Fig. 1. Overall workflow of the proposed supervised fine-tuned transformer for symbolic discovery of power system dynamics.

C. Dataset Generation Algorithms

The procedure for generating the training trajectories emulating power system three-phase-to-ground faults is summarized in Algorithm 1.

In the first step, we generate variants of two-machine and three-machine power systems with different network topologies and parameter settings. For each generated system, a power flow analysis is performed to verify the existence of a feasible steady-state operating point under the specified operating conditions. The resulting steady-state solution is then used to initialize the subsequent dynamic simulations, while cases without a feasible power flow solution are discarded.

Next, we simulate the trajectories of the generator phase angles during three-phase-to-ground fault under different fault locations and fault durations. The fault is assumed to occur on transmission lines close to the load buses. During dataset generation, all such fault locations are systematically explored. Then, different fault durations are selected under the same operation condition. The ODE systems of faulted and post-fault conditions are obtained and continuously integrated using the RK45 (Dormand–Prince) method.

Lastly, we gather the sample trajectories together with their symbolic expressions and split them to form the training, testing, and validation datasets for fine-tuning. Since trajectories generated from the same power system under different fault durations share the same underlying ODE system, the dataset is split at the system level rather than the trajectory level. This ensures that trajectories governed by the same ODE system do not appear across different splits, preventing data leakage and enabling a fair evaluation of model generalization.

IV. Numerical Studies

A. Test System and Dataset Generation

To investigate whether data from a different topology can compensate for limited training data, we construct two groups of power-system datasets based on 2-generator and 3-generator systems. For the 3-generator systems, we consider power networks consisting of 3 generators, 5 buses, and 6 transmission lines. A total of 12 network topologies are constructed. For each topology, the parameters listed in Table I are uniformly sampled within their prescribed ranges to generate multiple parameter sets. Specifically, each parameter category contains 10 sampled groups.

Algorithm 1 Data Generating

```
for each power system topology do
    for each set of generator, line, load parameters do
        if infeasible power flow case then
            Discard infeasible case
        end if
        for each fault location do
            for each fault duration do
                Simulate faulted and post-fault dynamics
                Collect the post-fault trajectory-ODE pairs
            end for
        end for
        Group trajectories of the same ODE system
        Split the grouped samples into different datasets
    end for
end for
```

TABLE I
Parameters contained in each generated parameter set

Category	Parameter	Description	Range (p.u.)
Generator	M	Inertia constant	[0.01, 0.05]
	x'_d	Transient reactance	[0.15, 0.35]
	P_m	Mechanical input power	[0.6, 2]
	V_{m0}	Initial voltage magnitude	[0.95, 1.05]
Line	R_l	Resistance	[0, 0.02]
	X_l	Reactance	[0.05, 0.2]
	y_c	Shunt conductance	[0.005, 0.02]
Load	p_f	Power factor (lagging)	[0.85, 0.95]
	ρ_L	Active-load distribution weight	[0.3, 1.0]

For load sampling, the active power loads are assigned only to load buses. For each load bus, a random distribution weight ρ_L is sampled, and all weights are normalized so that the total active load equals the total active power generation. The reactive load at each load bus is then computed from a lagging power factor p_f.

Since three-phase-to-ground fault is assumed to occur on the transmission lines close to the load buses, there are 4 to 6 possible fault locations for each topology depending on how the load buses are connected. Fault durations are selected from 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, and 0.20 s. The dataset size is determined by the combinations of 12 network topologies, 4 to 6 valid fault locations per topology, 9 fault durations, and the sampled parameter sets, where each parameter category contains 10 sampled groups. In total, 398,547 simulation trajectories are generated, with 80% samples for training, 10% samples for validation, and 10% samples for testing.

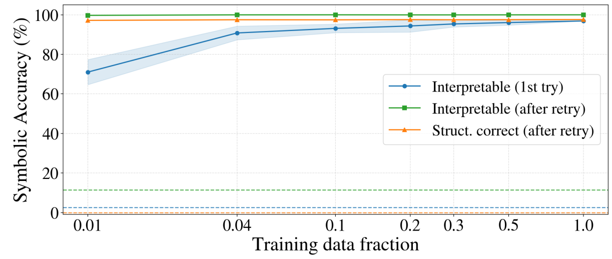
For the 2-generator systems, we consider two generators interconnected by two transmission lines. Therefore, only line and generator parameters are varied in the parameter generation process. A total of 90,000 simulation trajectories are generated, with 80% used for training, 10% used for validation, and 10% used for testing.

B. Evaluation Metrics

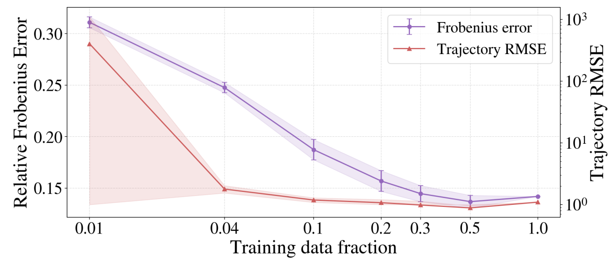
To quantify the performance of the proposed method, we employ the following evaluation metrics.

- *Trajectory Fidelity*: The root mean square error (RMSE) is calculated between the predicted state trajectories (rotor angles and speeds) and the ground truth obtained from numerical integration. Note that the predicted state trajectories are produced based on the predicted symbolic expression of the ODE system. The RMSE is given by $RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}||x_{true}(t) - x_{pred}(t)||_2^2}$, where $x_{true}(t)$ and $x_{pred}(t)$ denote the ground-truth and reconstructed state vectors at time step t.
- *Symbolic Accuracy*: For each test sample, we use two metrics to assess structural correctness and interpretability respectively. Structural correctness refers to whether the discovered ODE matches the ground truth physical swing equation, such as recovering the sinusoidal coupling terms, whereas interpretability only requires that the predicted tokens decode into a valid integrable differential equation. Accuracy is calculated as the ratio of samples with structurally correct or interpretable ODEs to the total number of samples.
- *Parameter Precision*: For correctly identified structures, we extract the coefficient matrices and compute the relative Frobenius norm error against the ground truth, defined as (10):

$$e_F(M_{pred}, M_{gt}) = \frac{\sqrt{\sum_{i,j}(M_{pred,ij} - M_{gt,ij})^2}}{\sqrt{\sum_{i,j} M_{gt,ij}^2}}, \quad (10)$$



(a) Symbolic accuracy. The dashed line indicates the accuracy of the pretrained ODEFormer.



(b) Trajectory fidelity and parameter prediction precision.

Fig. 2. Model performance under different training set sizes. The x-axis is plotted on a logarithmic scale.

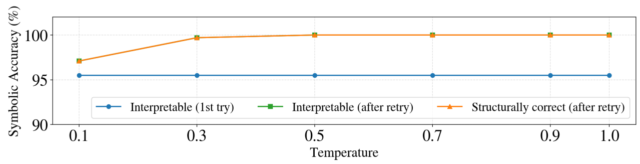
where M_{pred} and M_{gt} represent the predicted and ground truth coefficient matrices respectively. This metric provides a normalized, scale-invariant measure of the discrepancy between the predicted and actual parameters, effectively quantifying the model's precision in capturing the underlying grid dynamics at the parametric level.

C. Performance of the Supervised Fine-Tuned Transformer

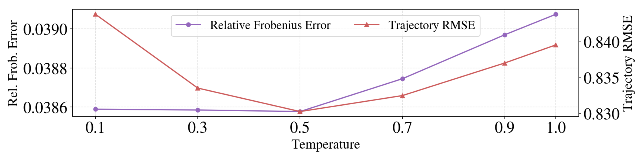
1) Effect of Training Set Size: To assess the effect of training dataset size, we trained a group of models using different fractions of the full training set and evaluated their performance under each setting. For each fraction, the training samples were randomly drawn from the 3-generator system training set according to the specified proportion. To reduce the impact of randomness in data sampling and model initialization, each training was repeated using five random seeds. Inferences are conducted using greedy decoding followed by beam sampling with temperature τ set to 0.7. The mean and variance of the evaluation metrics were reported in Fig. 2.

The experimental results show that, although the training set size affects interpretability in the first attempt, it does not influence structural correctness after five-seed retries. Both the trajectory RMSE and the parameter Frobenius norm error decrease as the training set size increases. However, increasing the training set size from 50% to 100% slightly worsens both metrics, suggesting that further enlarging the training set may even degrade model performance.

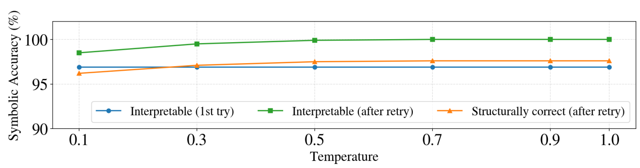
2) Effect of Beam Sampling Temperature: To test the influence of randomness on structural correctness, inferences are conducted with various beam sampling temperatures. We conduct experiments on both 2-generator and 3-generator samples. The result is reported in Fig. 3. The results indicate that a



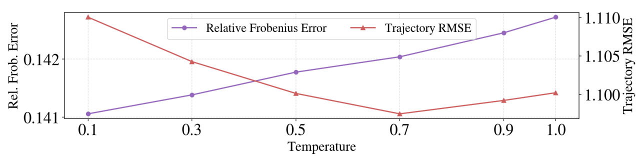
(a) Symbolic accuracy of 2-gen samples.



(b) Trajectory and parameter accuracy of 2-gen samples.



(c) Symbolic accuracy of 3-gen samples.



(d) Trajectory and parameter accuracy of 3-gen samples.

Fig. 3. Model performance with different temperature.

moderate increase in randomness improves the likelihood of discovering the correct structure, while excessive randomness leads to increased errors in trajectory and parameter prediction.

3) Cross-Topology Support under Limited Data: To examine whether data from another topology can compensate for insufficient training data, we train models on a mixed dataset consisting of 20% of the 3-generator samples and 20% of the 2-generator samples. The trained models are then evaluated separately on the 3-generator and 2-generator test sets. For comparison, we also train baseline models using single-topology datasets of the same size, i.e., only 20% of the 3-generator samples or only 20% of the 2-generator samples.

TABLE II
CROSS-TOPOLOGY PERFORMANCE

Train / Test	RMSE	Rel. Frob.	Struct. Corr.	Interp.
mix / 3-Gen	1.128 ±0.160	0.154 ±0.008	97.57 ±0.05	100.00 ±0.05
3-Gen / 3-Gen	1.075 ±0.138	0.156 ±0.010	97.56 ±0.05	99.96 ±0.05
mix / 2-Gen	1.374 ±0.137	0.055 ±0.008	99.90 ±0	99.90 ±0
2-Gen / 2-Gen	1.454 ±0.094	0.064 ±0.003	100.00 ±0	100.00 ±0

During inference, the temperature τ is set to 0.7 for 3-generator samples and to 0.5 for 2-generator samples. Each experiment was repeated using 3 random seeds, and the mean and variance of the evaluation metrics were reported in Table II. The results suggest that when training data are insufficient, higher-dimensional system data can improve the trajectory and parameter accuracy, while lower-dimensional system data can improve the structure correctness and interpretability.

V. Conclusion

In this paper, a supervised fine-tuned transformer framework based on ODEFormer is proposed for symbolic discovery of power system dynamics from post-fault transient trajectories. Numerical studies on multi-machine systems demonstrated that the model achieves high structural correctness and reliable parameter estimation under different training and inference settings. The results also indicate that when training data are insufficient, higher-dimensional system data can improve the trajectory and parameter accuracy, while lower-dimensional system data can improve the structure correctness and interpretability. Future work will extend the proposed framework to measurement-based datasets and more realistic large-scale system settings, including measurement noise, partial observability, parameter uncertainty. The underlying dynamic models will also be extended to incorporate damping, governor and excitation systems, inverter-based resource controls, and other relevant dynamic components, thereby improving the robustness and practical applicability of the framework.

References

[1] W. Cui, W. Yang, and B. Zhang, "A frequency domain approach to predict power system transients," *IEEE Trans. Power Syst.*, vol. 39, no. 1, pp. 465–477, 2024.

[2] M. Bossart, J. D. Lara, C. Roberts, R. Henriquez-Auba, D. S. Callaway, and B.-M. Hodge, "Acceleration of power system dynamic simulations using a deep equilibrium layer and neural ODE surrogate," *IEEE Trans. Energy Convers.*, vol. 40, no. 4, pp. 2710–2722, 2025.

[3] M. Karacelebi and J. L. Cremer, "Online neural dynamics forecasting for power system security," *Int. J. Electr. Power Energy Syst.*, vol. 167, p. 110566, 2025.

[4] X. Kong, K. Yamashita, B. Foggo, and N. Yu, "Dynamic parameter estimation with physics-based neural ordinary differential equations," in *2022 IEEE Power Energy Soc. Gen. Meeting (PESGM)*, 2022, pp. 1–5.

[5] A. Gorbunov, A. Dymarsky, and J. Bialek, "Estimation of parameters of a dynamic generator model from modal PMU measurements," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 53–62, 2020.

[6] S.-H. Lee, J.-H. Liu, B.-Y. Chen, and C.-C. Chu, "A two-stage data-driven method for estimating the system inertia utilizing event-driven pmu measurements," *IEEE Trans. Ind. Appl.*, vol. 59, no. 5, pp. 5243–5256, 2023.

[7] S. Zhang and N. Yu, "Learning power system dynamics with nearly-hamiltonian neural network," in *2023 IEEE Power Energy Soc. Gen. Meeting (PESGM)*, pp. 1–5.

[8] A. M. Stanković, A. A. Sarić, A. T. Sarić, and M. K. Transtrum, "Data-driven symbolic regression for identification of nonlinear dynamics in power systems," in *2020 IEEE Power Energy Soc. Gen. Meeting (PESGM)*, 2020, pp. 1–5.

[9] A. T. Sarić, A. A. Sarić, M. K. Transtrum, and A. M. Stanković, "Symbolic regression for data-driven dynamic model refinement in power systems," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2390–2402, 2021.

[10] S. d'Ascoli, S. Becker, A. Mathis, P. Schwaller, and N. Kilbertus, "ODEFormer: Symbolic regression of dynamical systems with transformers," 2023. [Online]. Available: https://arxiv.org/abs/2310.05573

[11] S. Becker, M. Klein, A. Neitz, G. Parascandolo, and N. Kilbertus, "Predicting ordinary differential equations with transformers," 2023. [Online]. Available: https://arxiv.org/abs/2307.12617