

Deep Learning-based Time Series Forecasting for Industrial Discrete Process Data

Olaf Sassnick^{*†‡}, Thomas Rosenstatter^{*†}, Andreas Unterweger^{†‡}, Stefan Huber^{*†}

^{*}Josef Ressel Centre for Intelligent and Secure Industrial Automation

[†]Salzburg University of Applied Sciences

[‡]Paris Lodron University of Salzburg

olaf.sassnick@fh-salzburg.ac.at

Abstract—With the introduction of Industry 4.0, the continuous collection and monitoring of industrial process data have become fundamental aspects of modern operational technology (OT) systems. The ability acquire high-resolution multivariate time series data, offers new opportunities for data-driven forecasting. Such forecasting facilitates proactive interventions, including process optimization and anomaly detection, with applications ranging from efficiency improvements to cybersecurity.

While deep learning-based forecasting models have demonstrated strong performance in various domains, their effectiveness for discrete manufacturing processes are insufficiently studied. In this paper, we introduce and publish a dataset that captures the key characteristics of discrete manufacturing time series data. We evaluate six state-of-the-art deep learning-based forecasting models on this dataset, identifying Crossformer and DUET as the best-performing approaches. However, we also identified that more research needs to be done for data generation, as it necessitates recursive forecasting due to real-time constraints. We already improved the performance significantly with DUET when performing a second-pass training based on the model's own forecasts.

Index Terms—forecasting, deep learning, dataset, discrete manufacturing

I. INTRODUCTION

With the establishment of Industry 4.0, the permeability of data [1] has become a key design principle of modern Operational Technology (OT) systems. Industrial process data is continuously collected and monitored. Advances in sensor technology, growing industrial cyber-physical systems (ICPS), onsite-computing resources, and high-throughput low-latency industrial networks allow for the acquisition of high-resolution data, capturing industrial processes in detail.

As a result, multivariate time series data with high resolution and dimensionality are potentially available. One promising approach for utilizing this information is data-driven forecasting, where insights from historical data allow to predict future outcomes. These predictions enable proactive interventions to improve the efficiency of an industrial process by either predicting the impact of changes of process parameters beforehand or detecting process anomalies and malfunctions

The financial support by the Christian Doppler Research Association, the Austrian Federal Ministry for Digital and Economic Affairs and the Federal State of Salzburg is gratefully acknowledged.

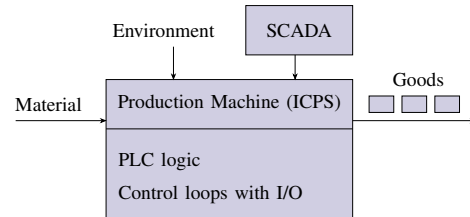


Figure 1. An ICPS performing a repetitive discrete task.

on-the-fly. For instance, time series forecasting can be used to optimize process parameters by predicting the impact of changes in advance. This capability can be leveraged to optimize product quality or enhance process efficiency, such as improving cost-effectiveness [2]. Another application is to detect process anomalies by identifying differences between the actual and predicted time series. Such anomalies may indicate system malfunctions or defects.

In this context, time series forecasting can strengthen cyber security as well, since detected process anomalies may result from cyber attacks [3]. Furthermore, data-driven forecasting can also be employed as generative models in honeypots, serving as defense mechanism by misleading potential attackers [4].

The recent advancements in artificial intelligence have resulted in a strong presence of AI-based forecasting models, with transformer architectures currently representing the state of the art [5]. The validation of universal forecasting models is carried out using an established set of openly available multi- and univariate time series from various domains [6]. Within this set, industrial processes are represented only by the energy industry, such as the Electricity Transformer Temperature (ETT) [7] or Electricity [8] datasets. However, the processes found in the energy industry are continuous, which are inherently different from those in the manufacturing industry, where piece-wise discrete items in repetitive procedures are produced. Compared to continuous processes, discrete manufacturing processes as illustrated in Figure 1 exhibit distinct characteristics, including higher process fidelity, greater determinism, and cyclic repetition (see Section II-A). Given these unique characteristics of time series found in discrete

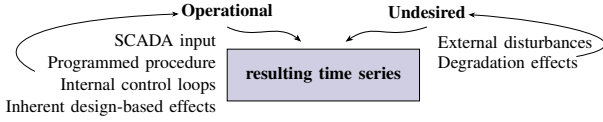


Figure 2. The different sources of a discrete manufacturing process acting in combination and affecting the result of the time series data.

manufacturing, our objective is to evaluate the performance of state-of-the-art deep learning based time series forecasting models for discrete manufacturing.

Contribution. First, we summarize the key characteristics of time series data found in discrete manufacturing processes and introduce an openly available dataset representing those characteristics. Consequently five deep learning based models are assessed for their effectiveness forecasting in discrete manufacturing time series, where it is also evaluated how these models respond to untrained scenarios, and test the generalization capability of each forecasting model.

While there is related work towards discrete manufacturing processes, such as production progress [9] and bottleneck [10] prediction, we have found no work similar to ours that investigates the data generation of multivariate time series of the discrete production process itself.

Outline. The remaining content is organized as follows: In Section II common discrete manufacturing process data characteristics are defined, and typical industrial forecasting applications are introduced. The research questions are discussed and the experiment procedure are outlined in Section III. Section IV describes the dataset and presents the procedure for each experiment. In Section V the results of the experiments are analyzed. Finally, the key findings are summarized, further implications are discussed, and future research directions are given in Section VI.

II. BACKGROUND

A. Characteristics of Discrete Manufacturing Process Time Series Data

Multivariate time series obtained from discrete manufacturing processes have characteristics which have so far not been considered in typical benchmark forecast datasets:

Process dynamics. Efficient production typically mandates high throughput, which in return requires high data sampling rates to capture fine-grained process details.

Deterministic behavior. Discrete manufacturing processes are typically subject to precise control with minimal fluctuations, requiring forecasting models to be sensitive to small deviations.

Cyclic repetition. In discrete manufacturing, the same process is repeated for each manufactured item. Modern Programmable Logic Controllers (PLCs), for instance, implement priority-based cyclic execution models [11]. This results in repeating temporal patterns, which appear periodically. In time series analysis, this characteristic often defined as seasonality [6].

Table I
OVERVIEW ON FORECASTING REQUIREMENTS OF APPLICATIONS

	AD	PO	DG	PM
Short-Term	•	•		
Long-Term				•
Real-Time	•	◦	•	
Detailing			•	
System Understanding		•	•	

Notes: AD: Anomaly Detection, PO: Process Optimization
DG: Data Generator (Honeyplot), PM: Predictive Maintenance

In the following, the focus is placed on the discrete manufacturing process of a single production machine, which can be considered an ICPS. As shown in Figure 2, the time series are the result of different sources acting in combination:

SCADA Input. Supervisory Control and Data Acquisition (SCADA) inputs that orchestrate and synchronize the individual ICPS, e.g., a start/stop signal.

Programmed procedure. Typically, the ICPS executes a pre-programmed procedure, such as a sequence of positions and orientations for the end effector of a robotic arm.

Internal Control Loops. Control loops within the ICPS regulate actuators according to sensor data to maintain a desired performance, e.g., temperature control for a heater element, or position control for an actuator [11].

Inherent Design-based Effects. Inherent due to the mechanical and electrical design of the ICPS, certain physical effects can be observed, which are expected as part of normal operation. For example, electrically, an analog signal can only be quantized with a certain precision, and voltage ripple caused by actuators might reduce the sensor readout precision. Mechanically, there can be geometrical tolerances, such as run-out imbalance or backlash in a reduction gear.

In addition to the previously listed sources, that are expected and part of normal operation, further undesired sources may contribute as well:

External Disturbances. Environmental or external factors that can influence the system's operation, such as abnormal temperature or unforeseen events, like an external force acting on a component of the ICPS.

Degradation Effects. The impact of wear-out of physical system components that can lead to performance degradation and, ultimately, failure over time.

B. Applications for Forecasting of Discrete Manufacturing Time Series Data

Depending on the type of application, different forecasting requirements are relevant, as summarized in Table I. For Anomaly Detection (AD) in Discrete Manufacturing, typically short-term forecasts are required, identifying unusual patterns or behavior that deviate from the forecast behavior. For example, in [12] the forging of metal components is monitored. Anomaly detection is typically performed online; as such, real-time capabilities are required, i.e., forecasting models need to be computationally efficient. The focus can be either on

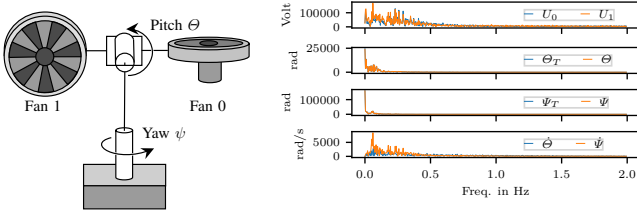


Figure 3. Left: The schematic representation of the CPS (from [4]) for the experiments with 2-DoF. Right: The frequency spectrum of the time series shown in Table II.

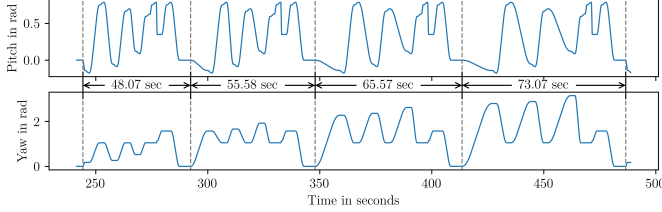


Figure 4. The cyclic procedure running on the CPS with yaw Ψ_T and pitch Θ_T changing over time (from [4]), with the system responding accordingly with its outputs listed in Table II.

machines or goods, and might as well be security-driven to detect ongoing attacks directed onto the industrial process.

In contrast, within Predictive Maintenance, indicating trends in long-term time series are used to forecast the Remaining Useful Life (RUL) of components. For instance, forecasts can be compared to existing Run-to-Failure Data, for example as it is commonly done for bearing components [13].

Process Optimization (PO) can be either online or offline, with online again mandating real-time capabilities and system understanding, as changes to the process are evaluated. For the last, the Data Generator (DG) represents the least common use-case. One application is a honeypot, where time series are generated by a forecasting model [4]. For this use case a stand-alone long-term data generation is required, meaning the forecasting model continuously generates forecasts based on own previous forecasts. As the goal of the honeypot is to make an intruder believe it is interacting with the real system, the resulting time series also needs to contain otherwise negligible characteristic details (indicated in Table I as Detailing), such as voltage ripple.

III. METHODOLOGY

This work aims to address two research questions (RQs):

RQ1. Which state-of-the-art deep learning based forecasting model performs well with discrete manufacturing process time series data?

RQ2. To what extent are deep learning based forecasting models capable of learning the underlying ICPS system dynamics, and forecasting untrained system responses?

Both, RQ1 and RQ2, require a suitable dataset as a foundation, representing a discrete manufacturing process. Therefore, as a first step, a dataset is designed to match the characteristics outlined in Section II-A. To address RQ1, models that perform

Table II
DESCRIPTION OF VARIABLES OF THE DATASET

Col.	Name	Description	Type	Sym.	Unit
0	Time	Elapsed time	—	t	s
1	Voltage0	Fan 0 Voltage	output	U_0	V
2	Voltage1	Fan 1 Voltage	output	U_1	V
3	Yaw	Actual yaw angle	measured	Ψ	rad
4	Pitch	Actual pitch angle	measured	Θ	rad
5	TargetYaw	Target yaw angle	input	Ψ_T	rad
6	TargetPitch	Target pitch angle	input	Θ_T	rad
7	YawDot	Yaw, angular velocity	estimated	$\dot{\Psi}$	rad/s
8	PitchDot	Pitch, angular vel.	estimated	$\dot{\Theta}$	rad/s
9	Override	Indicating an override of TargetPitch Θ_T	output	M	bool

well on the ETT [7] and Electricity [8] datasets - both representing physical processes - are selected. Additionally the traffic dataset is also considered, as it represents data with high seasonality. Each model's hyperparameters are tuned and finally, different performance aspects are evaluated on a test dataset.

In order to address RQ2, the dataset needs to be extended by introducing random changes to an input variable. We implement this by changing the input variable of the ICPS to a random value at a random timestep during the normal cyclic procedure. The response of the ICPS, which depends on both its current state and the assigned random value, is captured and compared to the forecast. This experiment is necessary as it indicates whether the forecasting model has learned the underlying system dynamics.

IV. EXPERIMENT SETUP

This section outlines the Dataset, Model Selection and Hyperparameter Tuning, and carried out Evaluations. Additionally, the dataset and full source-code to study and rerun the experiment is made available¹.

A. Dataset

The ICPS in this work is represented by a demonstrator consisting of a balancing beam with two mounted fans. It has two Degrees of Freedom (DoF), namely pitch Ψ and yaw Θ , as shown in Figure 3. By adjusting the airflow of the fans, the system can maintain configurable target angles for both yaw Ψ_T and pitch Θ_T . The datasets created with this CPS are modified from [4], in which the CPS is programmed to follow a cyclic procedure, resembling a typical pick-and-place process in discrete automation, as shown in Figure 4.

While the original datasets [4] contain 12 variables, this experiment is limited to 8 variables, excluding motor currents and speed trajectories. Additionally, the sampling rate is reduced to 100 Hz. As shown in the frequency spectrum plot in Figure 3, most of the time series have their main spectrum below 2 Hz. It mainly results in some information loss on the motor voltage 0 and 1 time series, where the reduction results

¹<https://github.com/JRC-ISIA/paper-2025-icps-discrete-manufacturing-time-series-forecasts>

Table III
RANKING OF FORECASTING MODELS BASED ON THE ETTM,
ELECTRICITY AND TRAFFIC DATASETS BASED ON [5]

Year	Model	ETTM	Electricity	Traffic
2025	DUET [5]	1	1	1
2024	PDF [14]	3	2	3
2023	DLinear [15]	5	6	7
2024	iTransformer [16]	8	4	6
2023	PatchTST [17]	5	8	5
2023	Crossformer [18]	11	4	5

in a Mean Absolute Error (MAE) of 0.005 V for each. Based on their frequency spectrum, the variables can be grouped into four clusters. The motor voltages exhibit the highest frequency components, followed by the angular velocities for yaw and pitch. Next, the actual yaw and pitch angles, and finally, the target yaw and pitch angles contain the lowest frequency components in the spectrum.

The second dataset is designed to address *RQ2*. While it initially follows the cyclic procedure illustrated in Figure 4, at a random time, the target pitch angle Θ_T is manually set to a random value between 0° and 40° . Shortly before this change, the override flag is activated. The system's response depends on whether it was stationary or in motion and on the distance to the previous Θ_T . The system holds Θ_T and Ψ_T for 8 seconds and afterwards restarts the cyclic procedure [4].

B. Model Selection and Hyperparameter Tuning

Based on the available results of 2025 in [5], the deep learning based forecast models are selected according to their ranking, as displayed in Table III.

The implementation of the forecasting models is used from [5] and [19]. The hyperparameter tuning is carried out using Ray [20] setup with the ASHA scheduler and the Hyperopt search algorithm. A batchsize of 32 is used, as proposed in [6].

The input sequence length (lookback or history) is set to 1600 values, and the horizon (lookahead) is set to 400 values. A sequence of 400 values is typically considered a long-term forecast, however with the high sampling rate of 100 Hz, this results in a 4-second forecast. While increasing the lookback slightly improves performance, it also significantly increases the model size.

Training is carried out with two data sources: one storing the normal cyclic procedure and a second storing the time series with random overrides (as described in Section IV-A), both joined into a single custom Torch Dataset. The train/validation/test split is set to 3/1/1 in both cases. For the normal cyclic data, training is performed using a rolling window strategy. For the random overrides, it is ensured that the override is always initiated within the lookback sequence, since otherwise, there are no indications for the override occurring, thus making it impossible for the model to forecast the system response.

C. Evaluation

For all evaluations, the MAE and Mean Squared Error (MSE) scores are used, where y represents the actual time series and $\hat{y}$ the forecast series, both of length n :

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad \text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

In order to allow for averaging of MAE and MSE with even weighting across all variables, each variable is normalized first with a standard scaler.

To answer *RQ1* and *RQ2*, four different evaluation scenarios are being studied:

Short-Term Forecasts: Initially, the focus is set on short-term forecasts, as they are relevant for online anomaly detection and process optimization. Given that the variables exhibit distinct frequency patterns, the performance is quantified per-variable with the MAE and MSE. This approach offers insight of the individual contributions of each variable to the overall error.

Downsampling of the Input Sequence: As Nie et al. [17] note, performance can improve with a longer lookback sequence. Given the high sampling rate as a data characteristic, one option is to lower the sampling rate of the input while retaining the sequence length, thus covering a longer time segment. For the forecast sequence, the decoder should be capable to learn maintaining the original sampling rate.

Recursive Forecasting for Data Generation: The third part evaluates the model with a recursive forecasting strategy, as this is necessary for a stand-alone data generator. With a recursive forecasting strategy, forecast sequences are used as part of the input sequence for subsequent forecasts. As already found in [4], it is expected that errors accumulate leading to a worsening MAE and MSE with each step of the recursion. In machine learning, training with lower-quality data is often used to improve model robustness. Based on this concept, a two-pass strategy is applied, where the model does a second training pass using its own previously generated forecasts.

System understanding with Override Conditions: The final part of the evaluation sets out to answer *RQ2* and addresses the model's capability to understand the system dynamics. In this scenario, the input sequence starts with normal cyclic behavior, but ends where an override takes place, with both an random target pitch and a corresponding override flag set at the end of the input sequence. Consequently, based on the trained override scenarios and normal operation time series data, the model needs to forecast the system behavior.

V. RESULTS AND ANALYSIS

The results for the **Short-Term Forecasts** are given in Table VI and visually exemplary results are displayed in Figure 6. This first evaluation scenario most closely resembles other typical forecasting tasks and, therefore, results in the best performance for each model. Despite the deterministic repeating cyclic procedure (Figure 4), iTransformer and DLinear already exhibit inferior forecasting results. Examining the

Table IV
MEAN RECURSIVE FORECASTING RESULTS OF THE OVERALL MAE
(AVERAGE ON ALL NORMALIZED VARIABLES) ACROSS $n = 100$ RUNS AND
UP TO 80 RECURSIONS.

Model/Iteration	0	20	40	60	80
DUET	0.046	0.339	0.597	0.844	0.921
DUET*	0.048	0.089	0.114	0.160	0.210
Crossformer	0.054	0.544	0.749	0.923	1.093
PDF	0.078	0.972	1.205	1.236	1.246
PatchTST	0.088	0.960	1.012	0.965	0.972
iTransformer	0.160	0.929	1.020	0.879	0.917
DLinear	0.626	1.082	1.039	1.052	1.074

* indicates a second training pass with own forecasts.

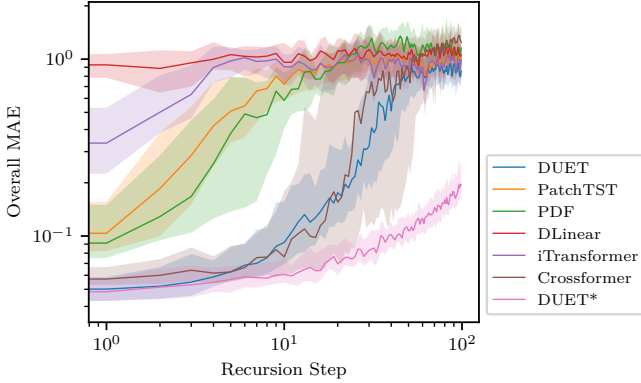


Figure 5. Recursive forecasting, showing the median as a plotted line with a shaded area representing the lower and upper quartiles for each model (* indicates a second training pass with own forecasts).

different variables, the highest MAE is present for the voltages, followed by the angular velocities, with the actual and target angles last. This correlates well with the frequency plot in Figure 3, where the voltages exhibit the highest frequency spectrum. The present voltage ripple is not predictable and, therefore, results in a higher MAE and MSE.

Table V contains results for Downsampling of the Input Sequence. While the results do not improve, the decoder of the model is capable of forecasting data with the original sampling rate. However, a slightly smaller standard deviation is visible, indicating that the twice-as-long lookback sequence makes the forecast more robust without compromising computational efficiency due to the unchanged parameter count of the model.

The results for **Recursive Forecasting for Data Generation** can be found in Table IV. Here, the potential use case as a stand-alone data generator is evaluated. After receiving an initial lookback input sequence of the multivariate time series, the forecast model recursively makes predictions, which are again based on its own previous forecasts. The real-time constraints demand that one cannot generate a single long-term forecast. As each forecast introduces some error, each model's performance degrades with every iteration step in recursive forecasting. The ranking among the models remains consistent: the model with the least error in single forecasts is best suited

Table V
AVERAGE FORECASTING RESULTS FOR $n = 100$ RUNS, EVALUATING THE
EFFECT OF 2X DOWNSAMPLING OF THE INPUT ('DS' INDICATING THE
DOWNSAMPLING).

	Average MAE and MSE	
DUET	0.049 ± 0.018	0.013 ± 0.027
DUETds	0.050 ± 0.017	0.012 ± 0.014

for the task. Generally, however, the degradation is severe, so none of the models seem suitable for a real application. The results, however, can be improved by applying a second-pass training, where the model is trained to make forecasts based on its own previous forecasts, as shown in Figure 5 and Table IV. In this case, the degradation is lower, with the MAE at times being over five times less.

Table VII shows the average results for **System Understanding with Override Conditions**, while Figure 7 gives two sample results. Both DUET and Crossformer perform well in this evaluation and also forecast the voltages with good accuracy, which typically exhibit highest frequency components (see Figure 3). As observable in Figure 7, offset errors in the prediction of yaw exist.

VI. CONCLUSION

Deep learning-based forecasting for multivariate time series has recently seen notable advancements. However, for the evaluated dataset, which resembles an industrial discrete process, significant deviations in forecasting quality remain.

In this paper, we identify key characteristics of discrete manufacturing time series data and introduce a publicly available dataset that exhibits these features. Our evaluation of state-of-the-art deep learning-based forecasting models on this dataset shows that Crossformer and DUET achieve the best performance.

However, we identified that more research needs to be done for data generation, as it necessitates recursive forecasting due to real-time constraints. We already improved the performance with a second-pass training based on the model's own forecasts. With this method, DUET showed significant performance improve compared to the other models. Additionally, an important aspect, though not included in this study, is the computational efficiency of each model.

REFERENCES

- [1] M. Hermann, T. Pentek, and B. Otto, "Design principles for Industrie 4.0 scenarios," in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, Koloa, HI, USA: IEEE, Jan. 2016, pp. 3928–3937. DOI: 10.1109/HICSS.2016.488.
- [2] D. Quesada, G. Valverde, P. Larrañaga, and C. Bielza, "Long-term forecasting of multivariate time series in industrial furnaces with dynamic Gaussian Bayesian networks," in *Engineering Applications of Artificial Intelligence*, vol. 103, p. 104301, Aug. 2021. DOI: 10.1016/j.engappai.2021.104301.
- [3] P. Filonov, A. Lavrentyev, and A. Vorontsov, *Multivariate industrial time series with cyber-attack simulation: Fault detection using an LSTM-based predictive data model*, arXiv:1612.06676, Dec. 2016. DOI: 10.48550/arXiv.1612.06676.

Table VI
FORECAST OF 400 LENGTH SEQUENCE FOR $n = 100$ RUNS

Model/Metric	Voltage U_0 MAE	Voltage U_1 MAE	$\dot{\Theta}$ MAE	$\dot{\Psi}$ MAE	Pitch Θ MAE	Yaw Ψ MAE	Target Θ_T MAE	Target Ψ_T MAE	Average MAE	MSE
DUET	0.128 ± 0.041	0.118 ± 0.037	0.098 ± 0.029	0.039 ± 0.018	0.010 ± 0.005	0.011 ± 0.007	0.009 ± 0.006	0.008 ± 0.006	0.047 ± 0.013	0.009 ± 0.008
Crossformer	0.132 ± 0.045	0.122 ± 0.047	0.108 ± 0.047	0.052 ± 0.039	0.024 ± 0.024	0.021 ± 0.015	0.025 ± 0.029	0.019 ± 0.014	0.057 ± 0.023	0.013 ± 0.017
PDF	0.150 ± 0.059	0.144 ± 0.059	0.153 ± 0.115	0.072 ± 0.057	0.079 ± 0.119	0.057 ± 0.072	0.067 ± 0.123	0.053 ± 0.078	0.086 ± 0.053	0.033 ± 0.058
PatchTST	0.158 ± 0.051	0.158 ± 0.084	0.144 ± 0.099	0.081 ± 0.058	0.075 ± 0.091	0.065 ± 0.069	0.060 ± 0.062	0.069 ± 0.068	0.090 ± 0.037	0.027 ± 0.033
iTransformer	0.335 ± 0.173	0.280 ± 0.149	0.285 ± 0.191	0.214 ± 0.144	0.105 ± 0.075	0.102 ± 0.080	0.111 ± 0.081	0.097 ± 0.080	0.170 ± 0.077	0.097 ± 0.113
DLinear	0.803 ± 0.495	0.690 ± 0.445	0.683 ± 0.403	0.634 ± 0.428	0.613 ± 0.374	0.518 ± 0.430	0.570 ± 0.352	0.524 ± 0.455	0.567 ± 0.243	0.732 ± 0.606

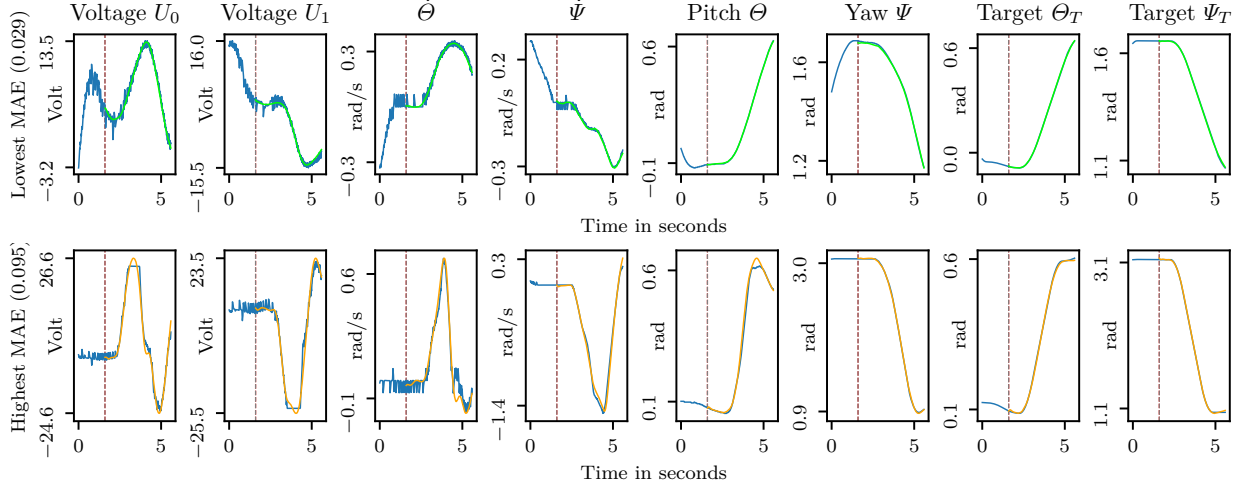


Figure 6. Lowest (first row) and highest (second row) average MAE result of the best forecasting model (DUET) for 8 variables of $n = 101$ generated forecasts.

Table VII
AVERAGE MAE AND MSE FOR $n = 100$ DIFFERENT OVERRIDE SCENARIOS

Model/Metric	MAE	MSE
DUET	0.059 ± 0.017	0.016 ± 0.012
Crossformer	0.074 ± 0.026	0.020 ± 0.015
PDF	0.278 ± 0.119	0.349 ± 0.299
PatchTST	0.314 ± 0.116	0.406 ± 0.303
iTransformer	0.351 ± 0.144	0.382 ± 0.302
DLinear	0.647 ± 0.256	1.069 ± 0.649

- [4] O. Saßnick, G. Schäfer, T. Rosenstatter, and S. Huber, “A generative model based honeypot for industrial OPC UA communication,” in *Computer Aided Systems Theory – EUROCAST 2024 – Extended Abstracts*, inpress, vol. 19, Feb. 2024. DOI: 10.48550/arXiv.2410.21574.
- [5] X. Qiu, X. Wu, Y. Lin, C. Guo, J. Hu, and B. Yang, “DUET: Dual clustering enhanced multivariate time series forecasting,” in *SIGKDD*, Accepted Version, 2025.
- [6] X. Qiu, J. Hu, L. Zhou, *et al.*, “TFB: Towards comprehensive and fair benchmarking of time series forecasting methods,” *Proc. VLDB Endow.*, vol. 17, no. 9, pp. 2363–2377, 2024.
- [7] H. Zhou, S. Zhang, J. Peng, *et al.*, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, pp. 11 106–11 115, May 2021. DOI: 10.1609/aaai.v35i12.17325.
- [8] A. Trindade, *ElectricityLoadDiagrams20112014*, 2015. DOI: 10.24432/C58C86.
- [9] W. Qian, Y. Guo, H. Zhang, *et al.*, “Digital twin driven production progress prediction for discrete manufacturing workshop,” *Robotics and Computer-Integrated Manufacturing*, vol. 80, p. 102 456, 2023. DOI: <https://doi.org/10.1016/j.rcim.2022.102456>.
- [10] D. Liu, Y. Guo, S. Huang, S. Wang, and T. Wu, “Dynamic production bottleneck prediction using a data-driven method in discrete manufacturing system,” *Advanced Engineering Informatics*, vol. 58, p. 102 162, 2023. DOI: <https://doi.org/10.1016/j.aei.2023.102162>.
- [11] M. A. Sehr, M. Lohstroh, M. Weber, *et al.*, “Programmable logic controllers in the context of Industry 4.0,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3523–3533, 2021. DOI: 10.1109/TII.2020.3007764.
- [12] B. Lindemann, N. Jazdi, and M. Weyrich, “Anomaly detection and prediction in discrete manufacturing based on cooperative LSTM networks,” in *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, ISSN: 2161-8089, Aug. 2020, pp. 1003–1010. DOI: 10.1109/CASE48305.2020.9216855.
- [13] B. Zhang, S. Zhang, and W. Li, “Bearing performance degradation assessment using long short-term memory recurrent network,” *Computers in Industry*, vol. 106, pp. 14–29, Apr. 2019. DOI: 10.1016/j.compind.2018.12.016.
- [14] T. Dai, B. Wu, P. Liu, *et al.*, “Periodicity decoupling framework for long-term series forecasting,” *International Conference on Learning Representations*, 2024.
- [15] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are Transformers Effective for Time Series Forecasting?” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 9, pp. 11 121–11 128, Jun. 2023. DOI: 10.1609/aaai.v37i9.26317.
- [16] Y. Liu, T. Hu, H. Zhang, *et al.*, “iTransformer: Inverted transformers are effective for time series forecasting,” *arXiv preprint arXiv:2310.06625*, 2024.
- [17] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, “A time series is worth 64 words: Long-term forecasting with transformers,” in *International Conference on Learning Representations*, 2023.
- [18] Y. Zhang and J. Yan, “Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting,” in *International Conference on Learning Representations*, 2023.
- [19] Y. Wang, H. Wu, J. Dong, Y. Liu, M. Long, and J. Wang, “Deep time series models: A comprehensive survey and benchmark,” 2024. DOI: 10.48550/arXiv.2407.13278.

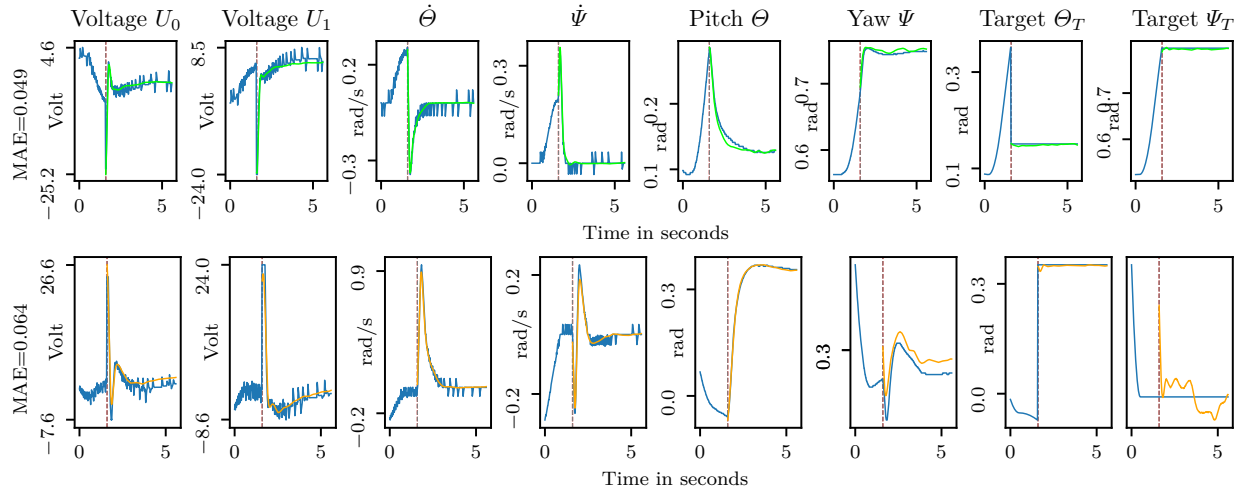


Figure 7. Two sample results of the best forecasting model (DUET) for 8 variables, where the system response to the manual change of the target pitch is forecast.

- [20] J. Bergstra, D. Yamins, and D. Cox, “Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures,” in *Proceedings of the 30th International Con-*

ference on Machine Learning, S. Dasgupta and D. McAllester, Eds., ser. Proceedings of Machine Learning Research, vol. 28, Atlanta, Georgia, USA: PMLR, 2013, pp. 115–123.