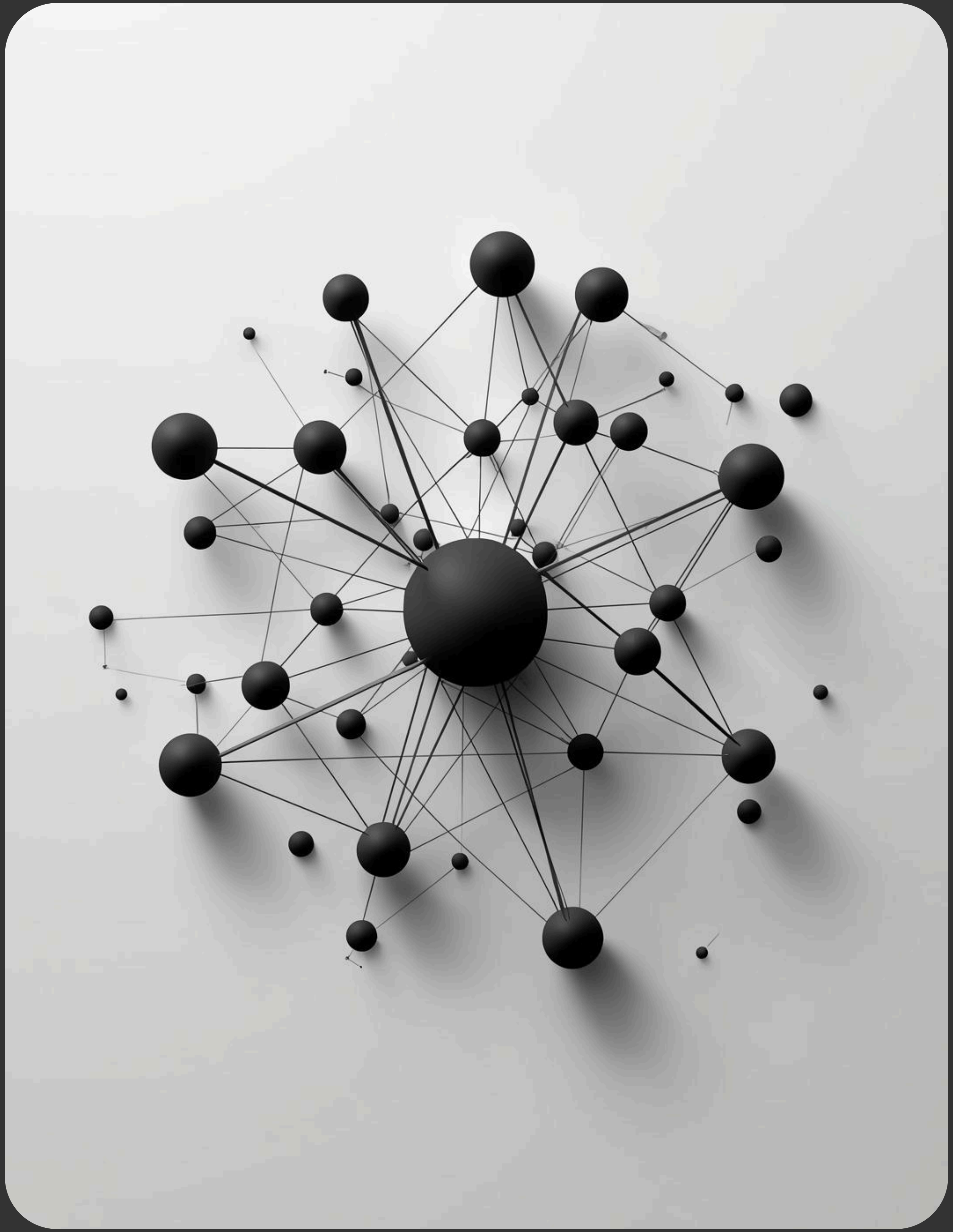


CH512: Machine Learning

MACHINE LEARNING BASED SOFT-SENSOR MODELING FOR SULPHUR RECOVERY UNIT (SRU)

Course Project Part 1: Function approximation methods

Course Project Part 2: Nonlinear function approximation techniques

SUBMITTED TO:
Dr. JAYARAM VALLURU

SUBMITTED BY:
NAME: AYATI BOHRA
ENTRY NO: 2023CHB1064

Introduction

- The Sulphur Recovery System (SRU) is a critical refinery/gas-processing unit that removes hydrogen sulfide (H_2S) from sour gas and converts it into elemental sulphur.
- It ensures compliance with emission limits, prevents corrosion, and produces saleable sulphur.
- Works based on the Modified Claus Process.

SULFUR RECOVERY UNIT (SRU)

SRU follows the Modified Claus Process in two stages:

◆ Thermal Stage :

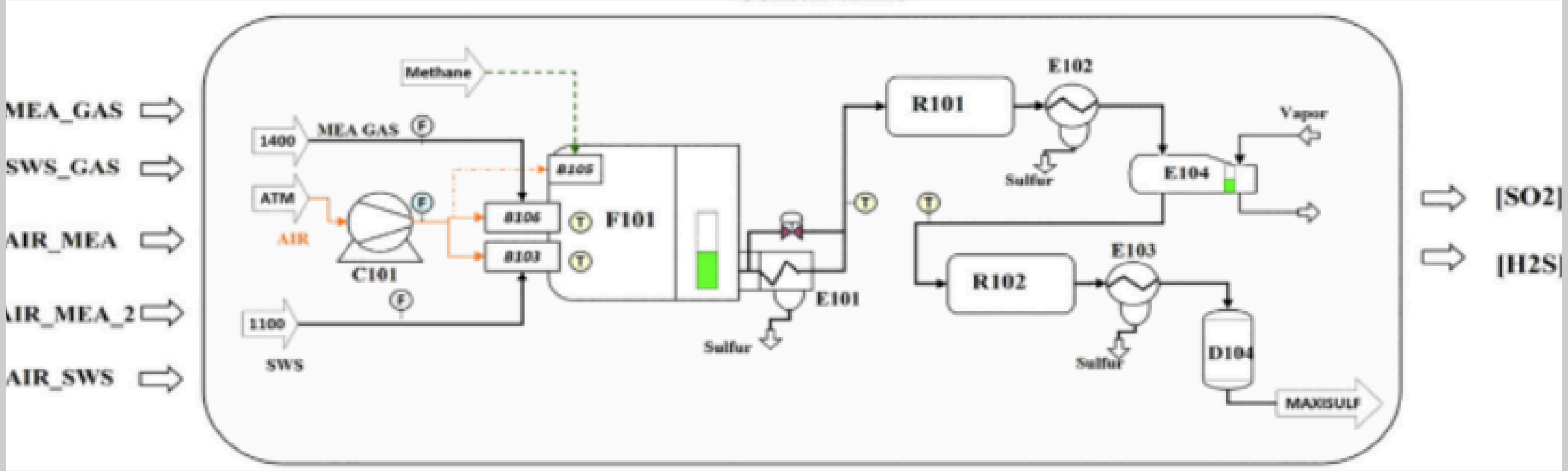
In the thermal stage, about one-third of hydrogen sulfide (H_2S) is partially combusted with air to produce sulfur dioxide (SO_2) at very high temperatures.
This step supplies the heat required for the process and removes unwanted impurities from the gas stream.

◆ Catalytic Stage :

In the catalytic stage, the remaining H_2S reacts with the SO_2 over catalyst beds to form elemental sulfur and water vapor.
Lower temperature operation in this stage favors higher sulfur recovery and efficient conversion.

To get maximum recovery, the plant must strictly maintain the H_2S : SO_2 ratio. Any imbalance causes emissions and loss of sulfur yield.

SRU DIAGRAM



Code	Process Variable	Meaning in SRU
IN1	**MEA_GAS**	Acid gas from amine unit
IN2	**SWS_GAS**	Sour water stripper gas
IN3	**AIR_SWS**	Combustion air to SWS gas
IN4	**AIR_MEA**	Combustion air to MEA gas
IN5	**UR_MEA**	MEA gas flow to thermal reactor

Output	Component
Out1	**H_2S**
Out2	**Sulfur**

MAJOR CHALLENGES WITH DIRECT ANALYZERS:

- Very expensive to install and maintain
- Slow response time due to sampling and analysis delays
- Easily damaged in corrosive or harsh environments
- Require frequent calibration and maintenance

PROBLEMS FACED BY OPERATORS

- No instant feedback on process performance
- Lack of continuous, online product quality measurement
- Unable to predict failures or abnormal conditions early
- Slow decision-making due to delayed data

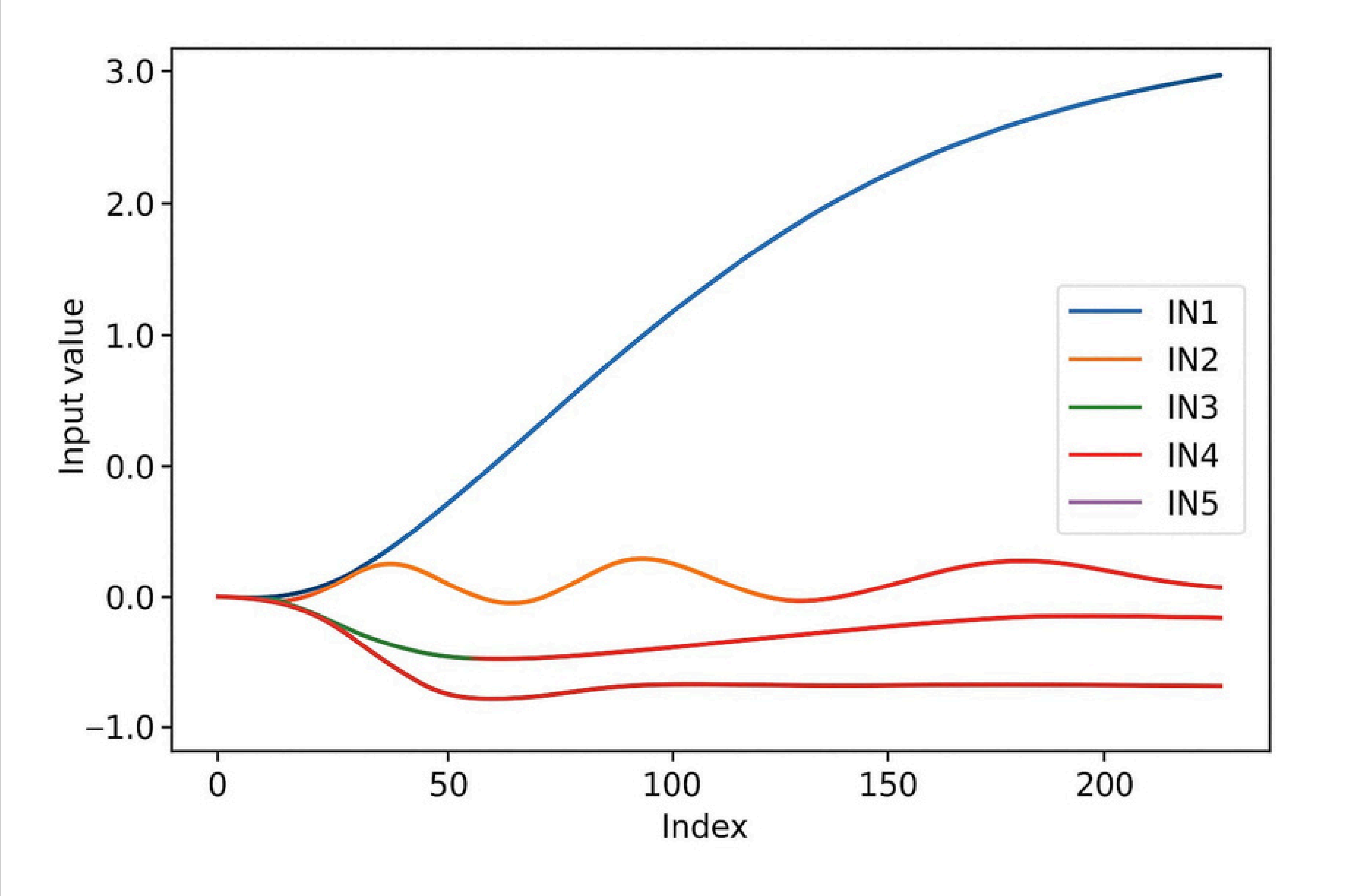
SOLUTION: SOFT SENSORS

- Estimate key outputs using easily measurable inputs
- Use machine learning models for accurate predictions
- Provide real-time monitoring and faster response
- Offer early fault detection and warning alerts
- Reduce dependency on physical analyzers
- Lower operational and maintenance costs

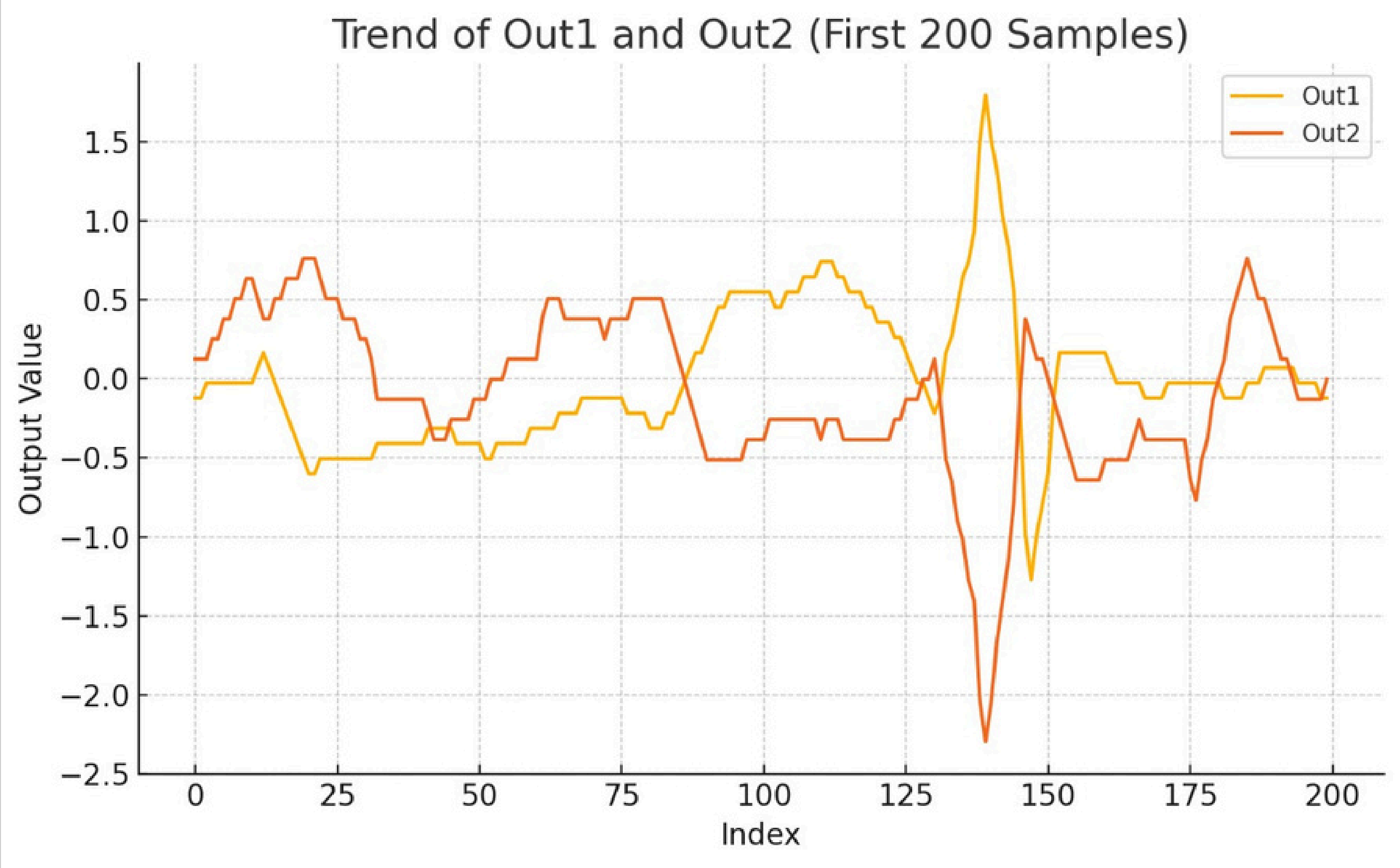
ANALYSIS OF INPUT TREND: All inputs except IN2 show a smooth increasing trend, with IN1 and IN4 rising the most strongly. IN2 oscillates with mild ups and downs, while IN5 stays almost constant with very small fluctuations.

ANALYSIS OF OUTPUT TREND: The first 200 samples show moderate oscillations in both outputs with small fluctuations around zero and occasional sharper peaks and dips indicating dynamic process behavior. Overall, Out1 and Out2 follow similar patterns, capturing the typical noise and variability of SRU tail-gas concentrations.

INPUT TRENDS



OUTPUT TRENDS



COURSE PROJECT PART 1

1. BASIC PIPELINE:

The Input(IN1, IN2, IN3, IN4, IN5) and Output (OUT1, OUT2) were loaded and split into train (70%) and test (30%). These scaled or selected features in multiple modeling approaches.

METHOD 1 – CORRELATION-BASED FEATURE SELECTION + LINEAR REGRESSION IDEA:

OUTPUT 1

Rank	Input	Correlation
1	IN1	0.6248
2	IN3	0.5722
3	IN5	0.4287
4	IN2	0.3883
5	IN4	0.3011

OUTPUT 2

Rank	Input	Correlation
1	IN3	0.4151
2	IN1	0.3819
3	IN4	0.2037
4	IN2	0.2016
5	IN5	0.1018

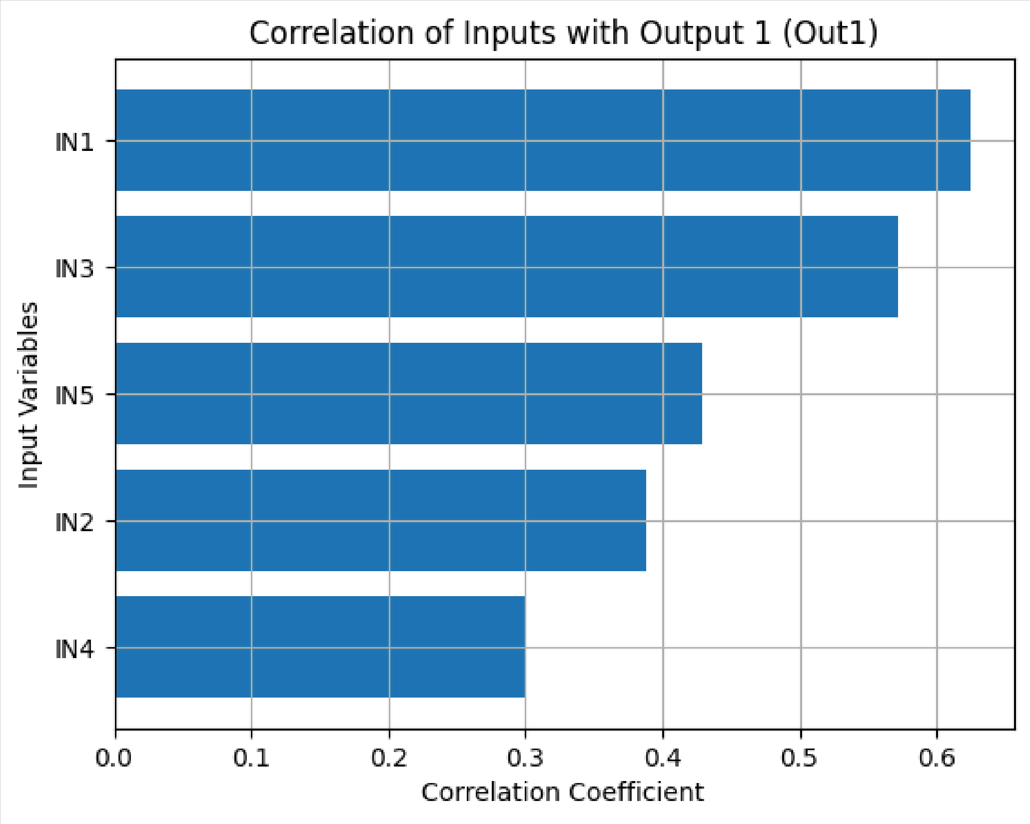
Output	Most Influential Inputs
Out1	IN1 → IN3 → IN5 → IN2 → IN4
Out2	IN3 → IN1 → IN4 → IN2 → IN5

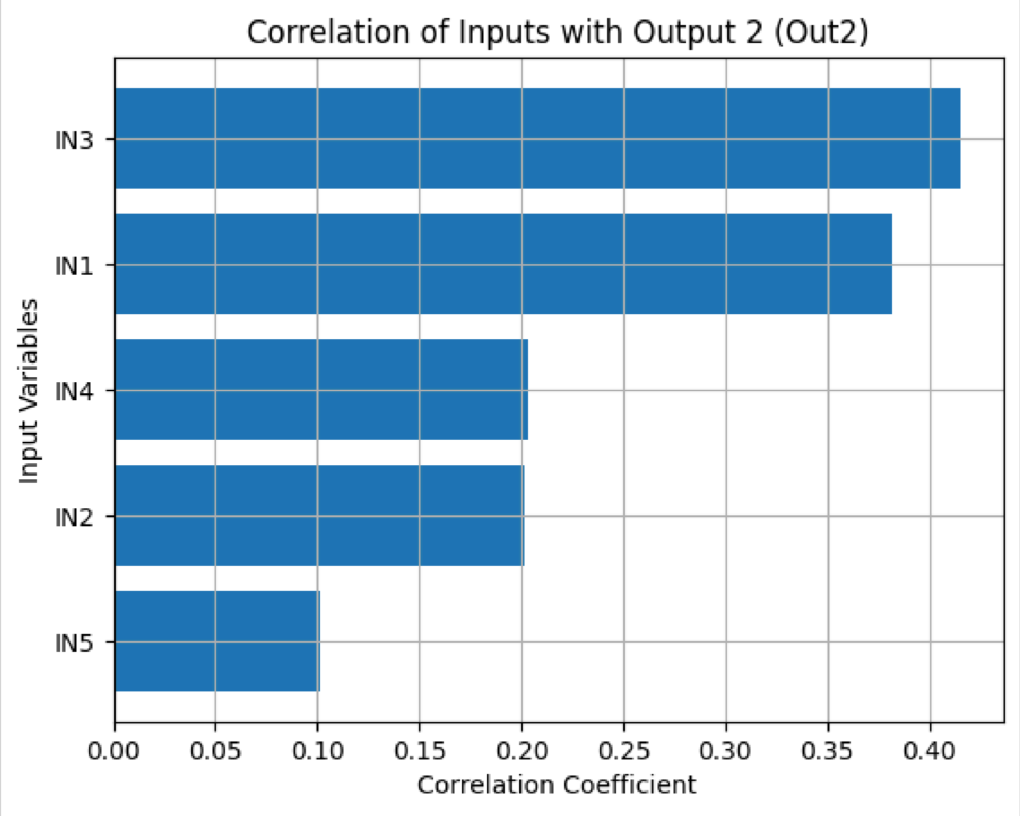
OUTPUT 1:

- IN1 and IN3 dominate Output 1, both having strong correlation (> 0.5).
- IN5 and IN2 show moderate effect, refining the output.
- IN4 has the weakest influence.

OUTPUT 2:

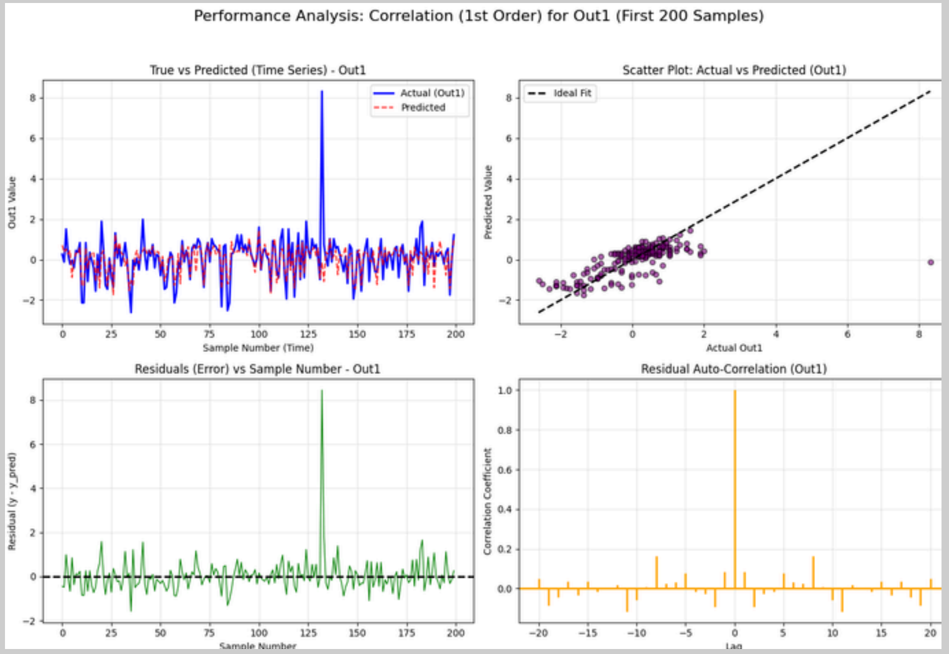
- IN3 is the primary controller for Out2.
- IN1 contributes significantly but weaker than for Out1.
- IN4 & IN2 provide mild influence.
- IN5 has very little effect.

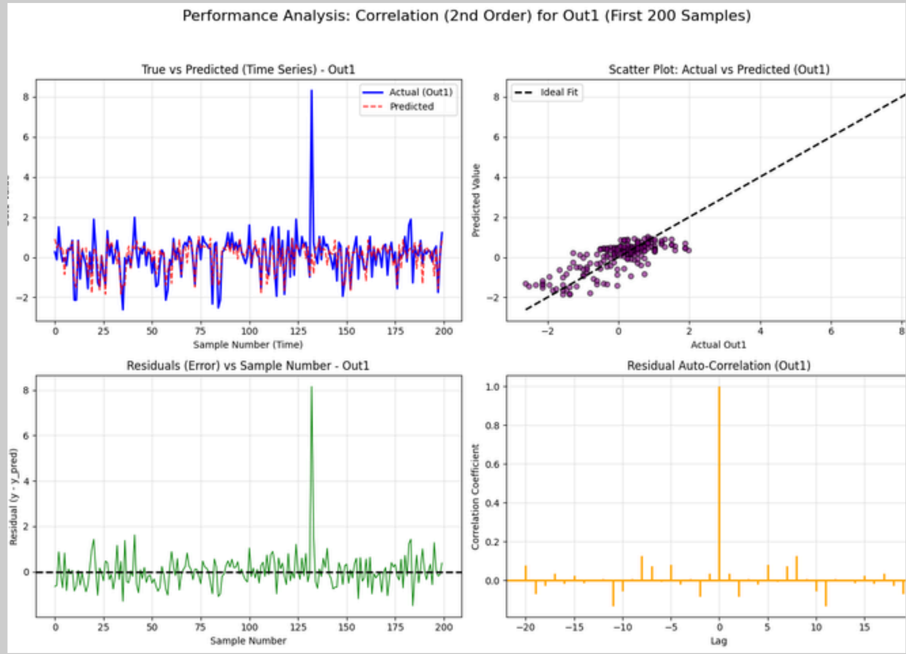




CORRELATION-BASED REGRESSION MODELS FOR OUTPUT 1

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Correlation (1st Order)	0.4285	0.4598	0.7348	2332.79	-2651.54	-2613.31
Correlation (2nd Order)	0.4774	0.5039	0.7041	2142.43	-2989.39	-2855.59

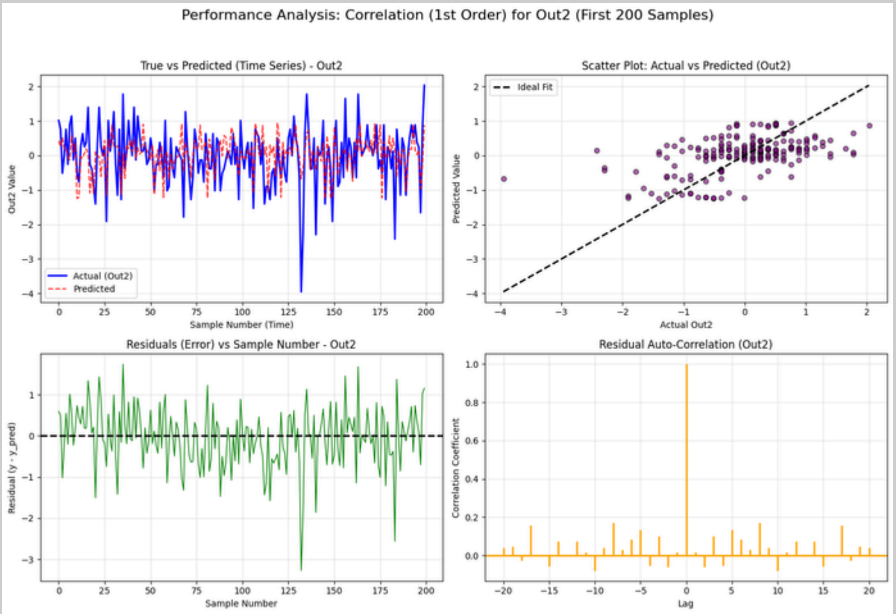


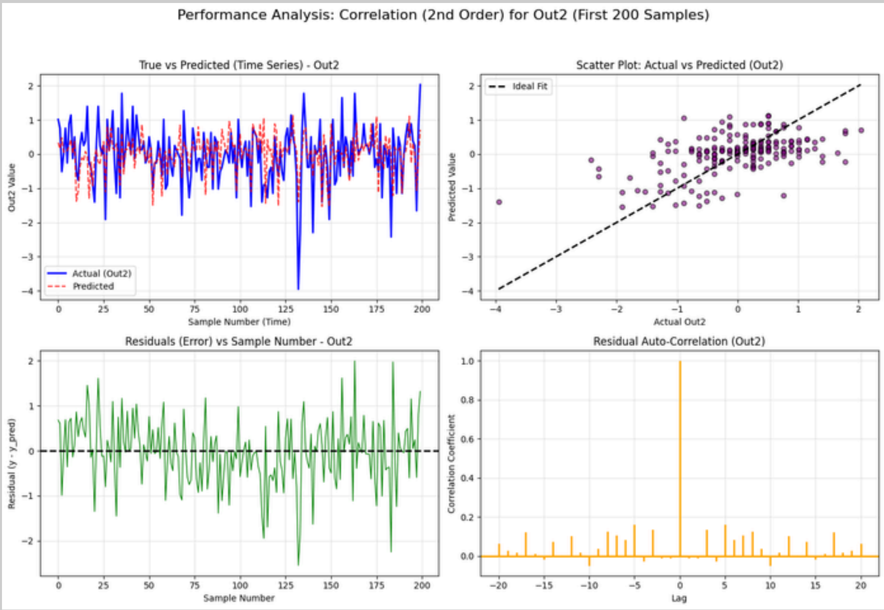


ANALYSIS: Out1 displays nonlinear behavior that the first-order model cannot fully capture, while second-order regression improves fit. The second-order model shows significantly lower AIC and BIC which means that complexity added by polynomial terms is justified.

CORRELATION-BASED REGRESSION MODELS FOR OUTPUT 2

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Correlation (1st Order)	0.217	0.2124	0.8691	3263.69	-1200.6	-1162.37
Correlation (2nd Order)	0.257	0.255	0.8452	3087.08	-1410.99	-1277.19

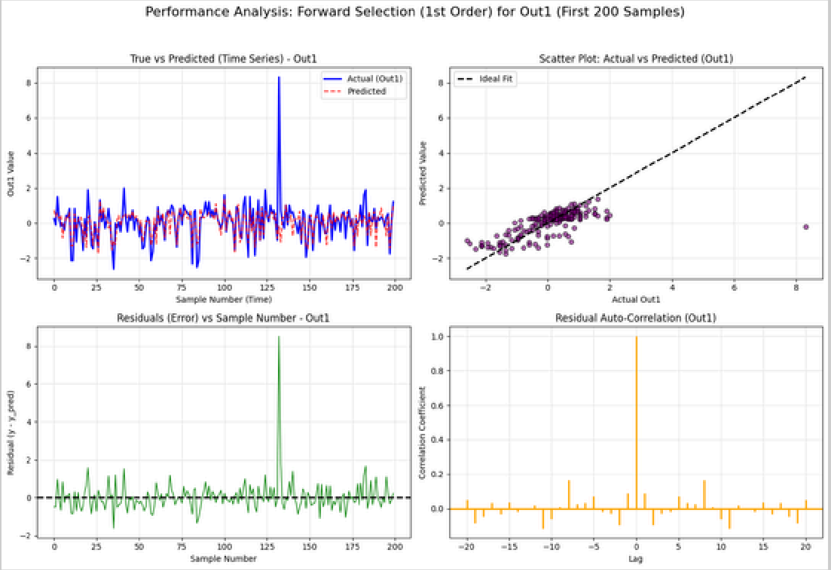


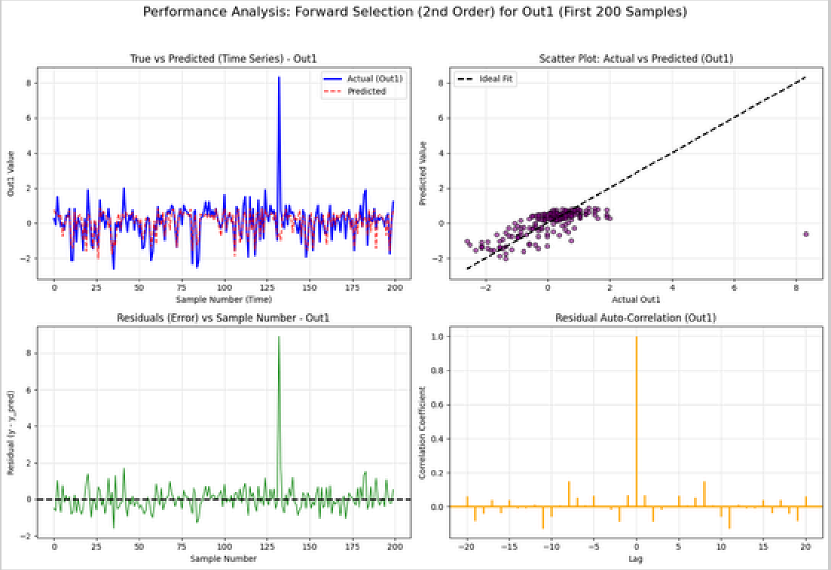


ANALYSIS: For Out2, both models show relatively low predictive performance compared to Out1. This implies that: Out2 likely depends on time dynamics. The second-order model has much lower AIC and BIC than first- order model.

FORWARD SELECTION REGRESSION MODELS FOR OUTPUT 1

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Forward Selection (1st Order)	0.4278	0.4591	0.7352	2335.8	--2647.98	-2616.13
Forward Selection (2nd Order)	0.4515	0.4749	0.7244	2267.44	-2756.34	-2660.77

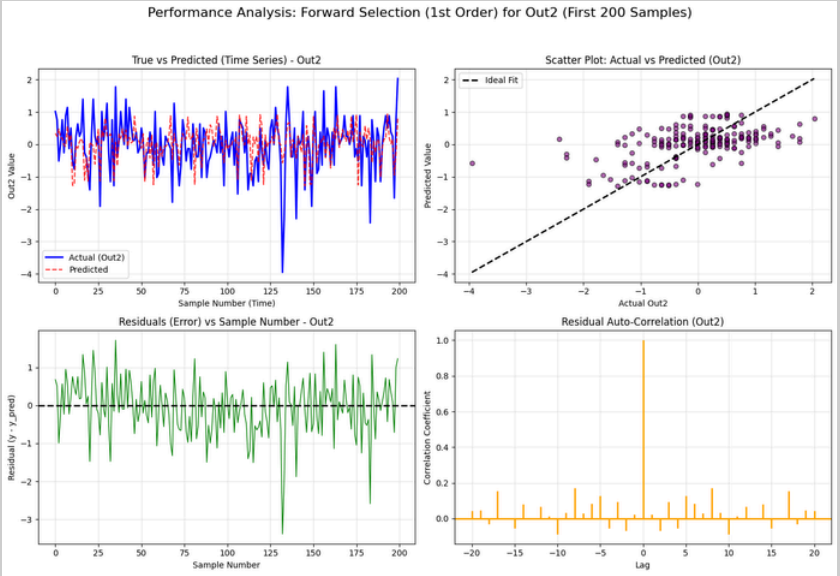




ANALYSIS: The second-order model improves prediction accuracy by increasing test R^2 and reducing RMSE and SSE. Lower AIC and BIC confirm that Out1 has nonlinear behavior and is better captured by polynomial modeling.

FORWARD SELECTION REGRESSION MODELS FOR OUTPUT 2

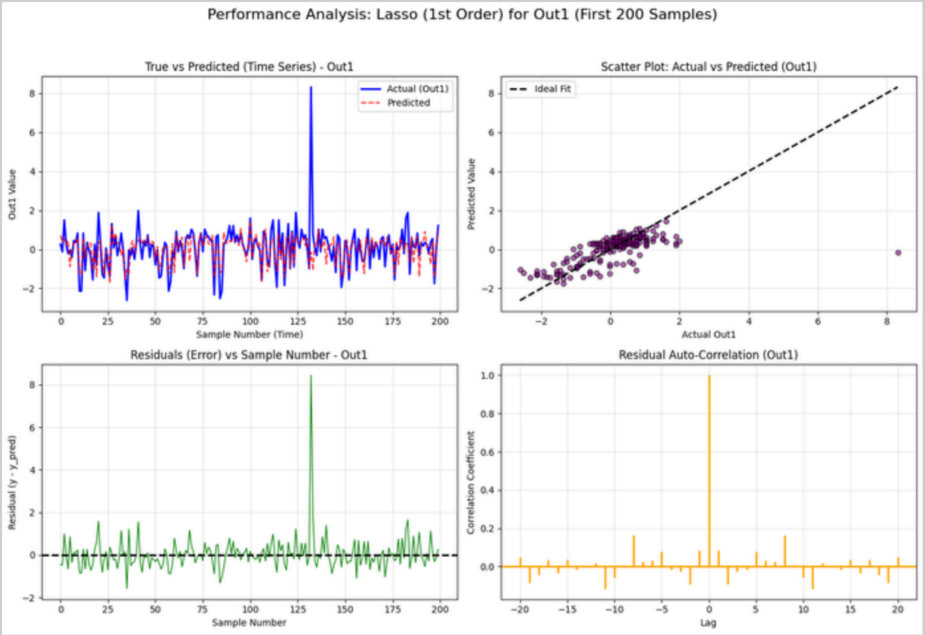
Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Forward Selection (1st Order)	0.2159	0.211	0.8699	3269.49	-1194.93	-1163.07
Forward Selection (2nd Order)	0.2468	0.2427	0.8522	3138	-1352.3	-1256.73

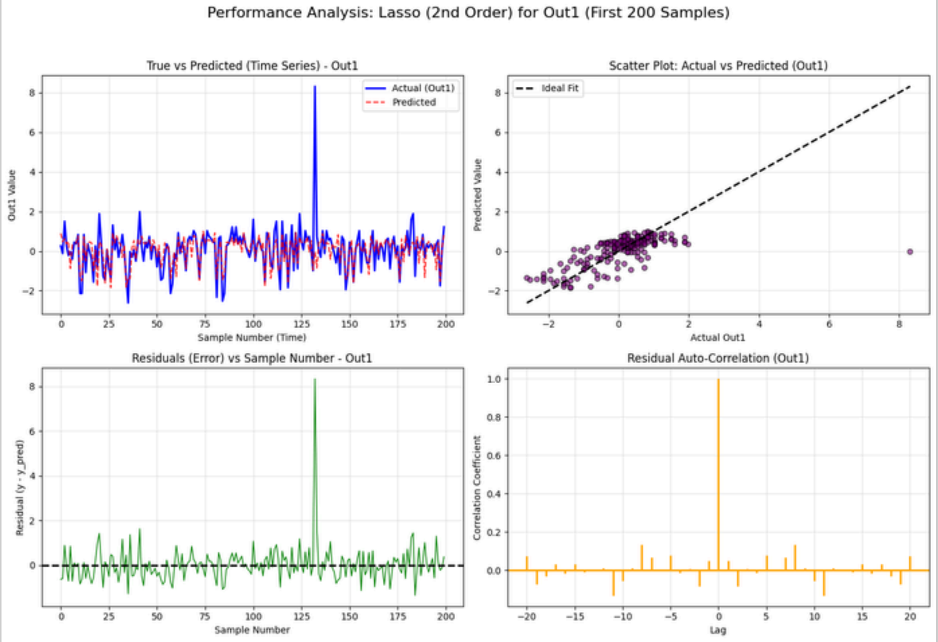


ANALYSIS: The second-order model shows only minor improvement for Out2.
This suggests Out2 requires dynamic models like LSTM for accurate prediction.

LASSO REGRESSION MODELS FOR OUTPUT 1

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
LASSO (1st Order)	0.4285	0.4598	0.7348	2332.92	-2651.33	-2613.1
LASSO (2nd Order)	0.475	0.5019	0.7055	2150.79	-2972.56	-2838.76

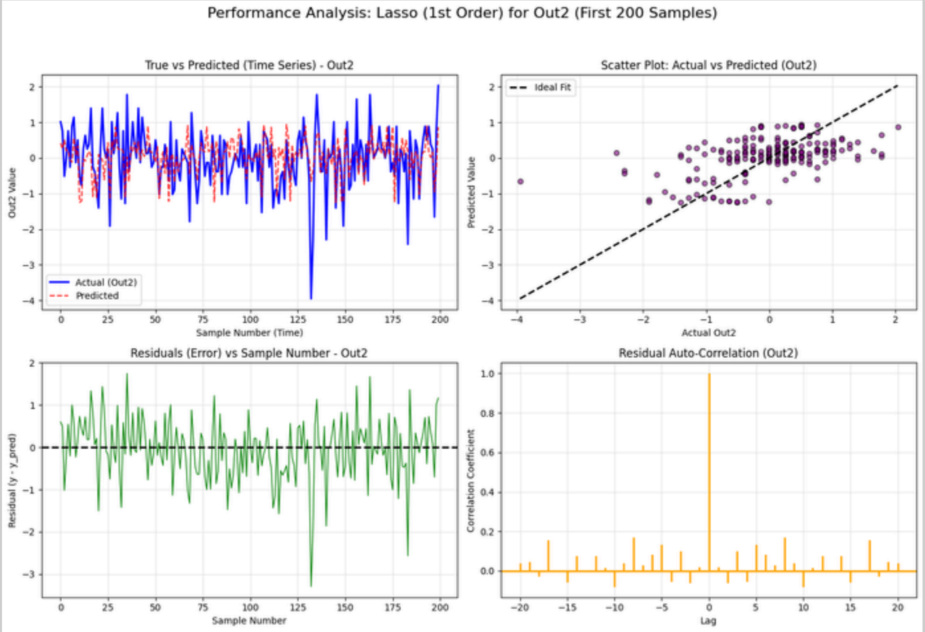


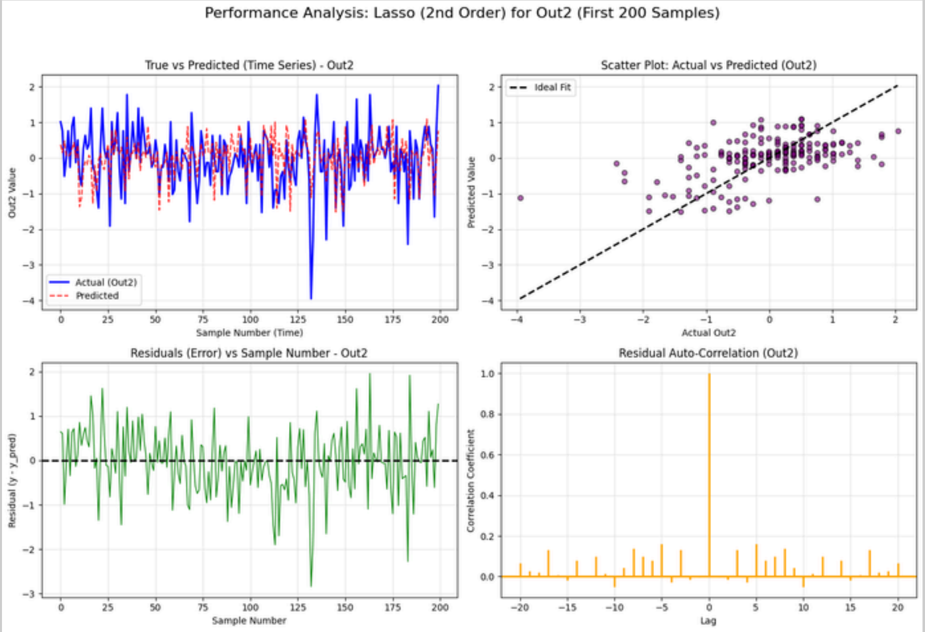


ANALYSIS: The second-order LASSO model improves accuracy by increasing test R^2 and reducing error values. Lower AIC and BIC confirm that Out1 is nonlinear and benefits from regularized polynomial modeling.

LASSO REGRESSION MODELS FOR OUTPUT 2

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
LASSO (1st Order)	0.217	0.2124	0.869	3263.38	-1201.01	-1162.78
LASSO (2nd Order)	0.2565	0.2543	0.8456	3089.95	-1406.97	-1273.18

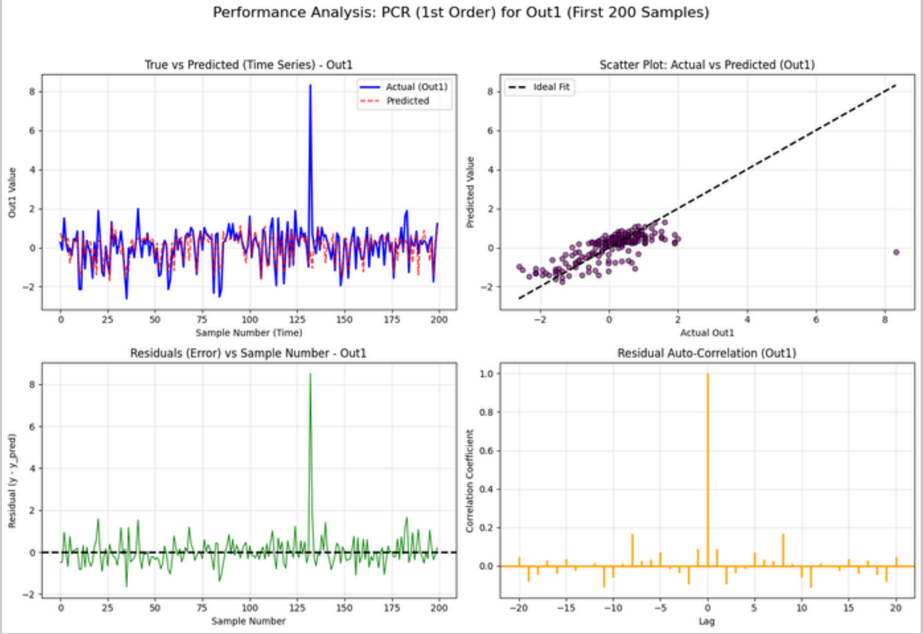


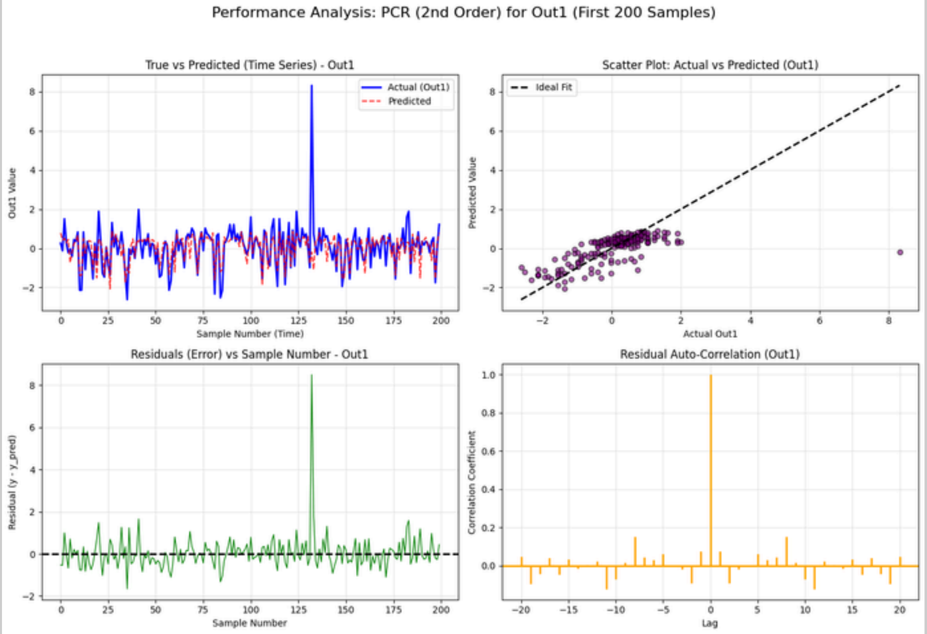


ANALYSIS: The second-order LASSO model gives only a slight improvement for Out2, with test R^2 remaining low. This confirms that Out2 requires dynamic models like LSTM for accurate prediction.

PCR MODELS FOR OUTPUT 1 (OUT1)

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
PCR (1st Order)	0.4248	0.4559	0.7374	2349.77	-2624.23	-2598.74
PCR (2nd Order)	0.4374	0.4632	0.7324	2317.88	-2671.27	-2607.56

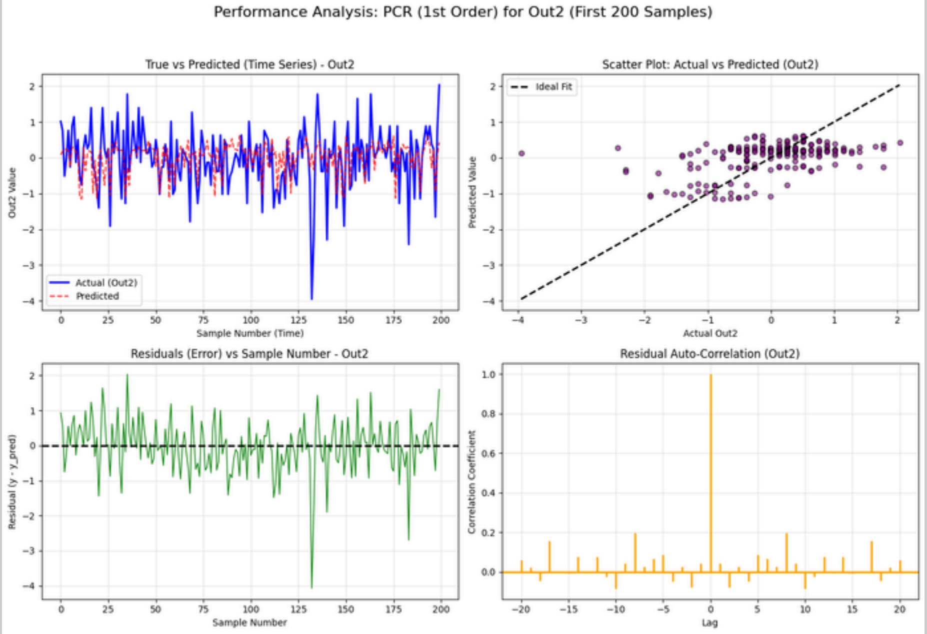




ANALYSIS: The second-order PCR model slightly improves prediction accuracy for Out1 compared to the first-order model. However, PCA reduces important information, making PCR less effective than feature-based regression.

PCR MODELS FOR OUTPUT 2 (Out2)

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
PCR (1st Order)	0.1749	0.1705	0.8919	3437.19	-980.79	-955.3
PCR (2nd Order)	0.1882	0.182	0.8857	3389.54	-1029.11	-965.4



ANALYSIS: PCR performs poorly for Out2 even after adding polynomial terms, with very low test R^2. This shows Out2 is dynamic and cannot be captured by static regression or PCA-based models.

RIDGE REGRESSION MODELS FOR OUTPUT 1

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Ridge Regression	0.4285	0.4598	0.7348	2332.83	-2651.49	-2613.27

ANALYSIS: Ridge regression gives similar performance to basic linear models with no significant improvement. This suggests that regularization alone cannot capture nonlinearity in Out1.

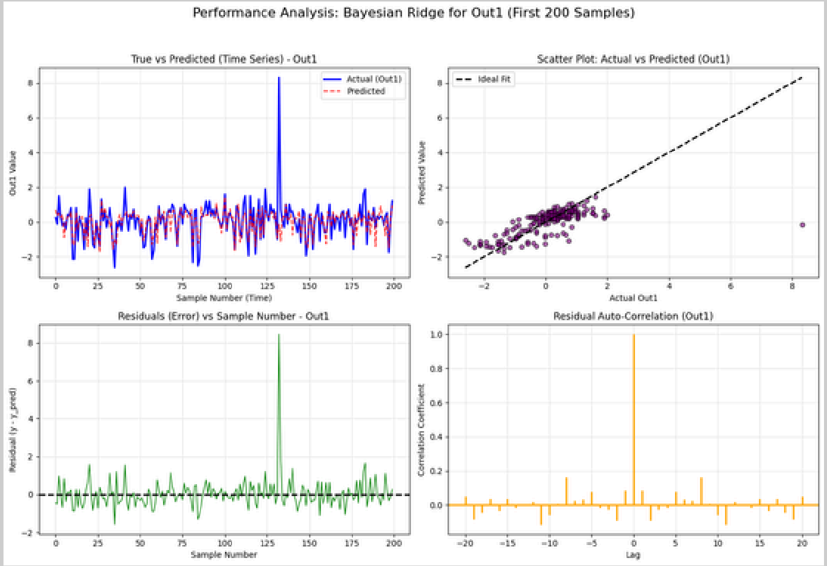
RIDGE REGRESSION MODELS FOR OUTPUT 2

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Ridge Regression	0.217	0.2124	0.8691	3263.62	-1200.69	-1162.46

ANALYSIS: Ridge regression performs poorly for Out2 with very low test R^2.
Static regularization models are insufficient to describe Out2 dynamics.

BAYESIAN RIDGE REGRESSION MODELS FOR OUTPUT 1

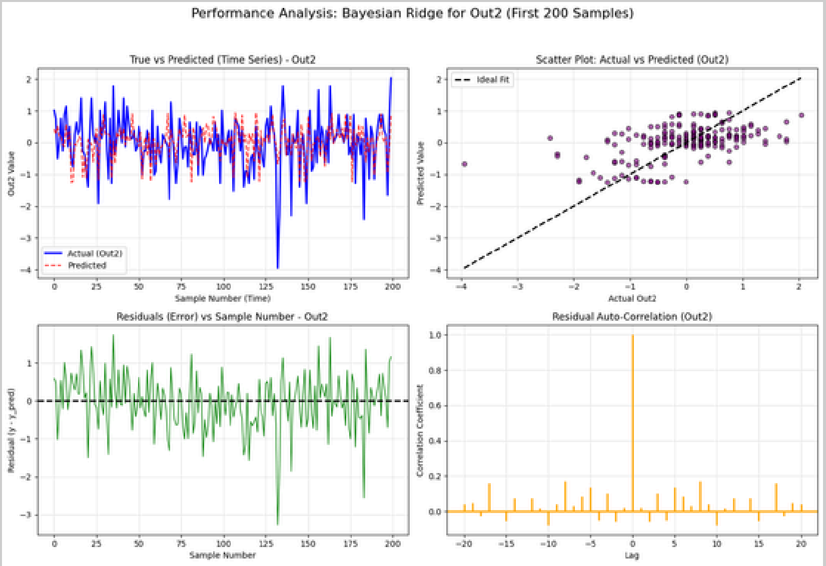
Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Bayesian Ridge	0.4285	0.4598	0.7348	2332.84	-2651.47	-2613.25



ANALYSIS: Bayesian Ridge offers no improvement over Ridge or basic linear models for Out1. Regularization alone is insufficient to capture nonlinear effects in the process.

BAYESIAN RIDGE REGRESSION MODELS FOR OUTPUT 2

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Bayesian Ridge	0.217	0.2124	0.8691	3263.52	-1200.82	-1162.59



ANALYSIS: Bayesian Ridge performs poorly for Out2 with very low test accuracy. Dynamic models are required to represent time-based system behavior.

COURSE PROJECT PART 2

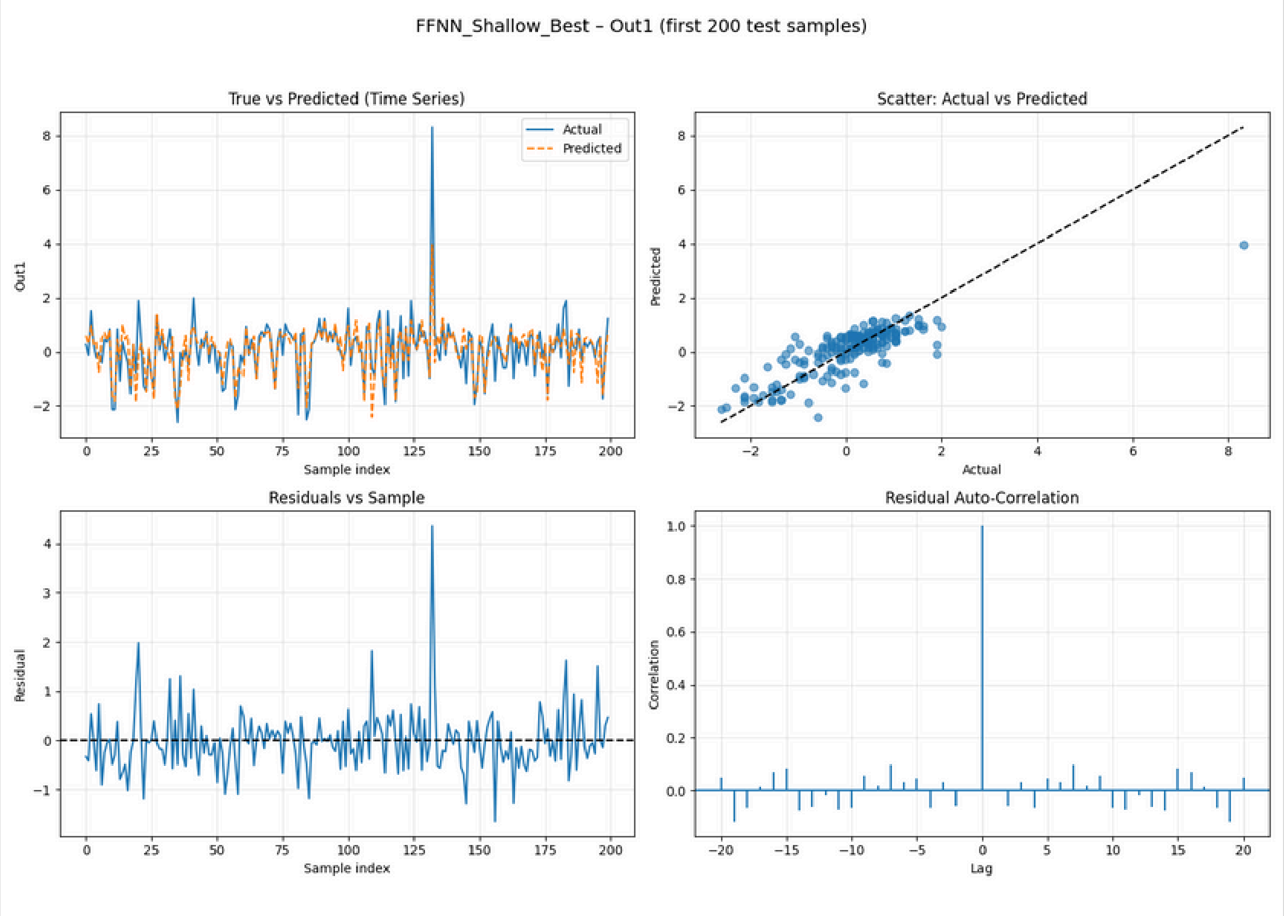
HYPERPARAMETER TUNING (BEST MODELS):

BEST FEEDFORWARD NEURAL NETWORK CONFIGURATIONS FOR OUT1

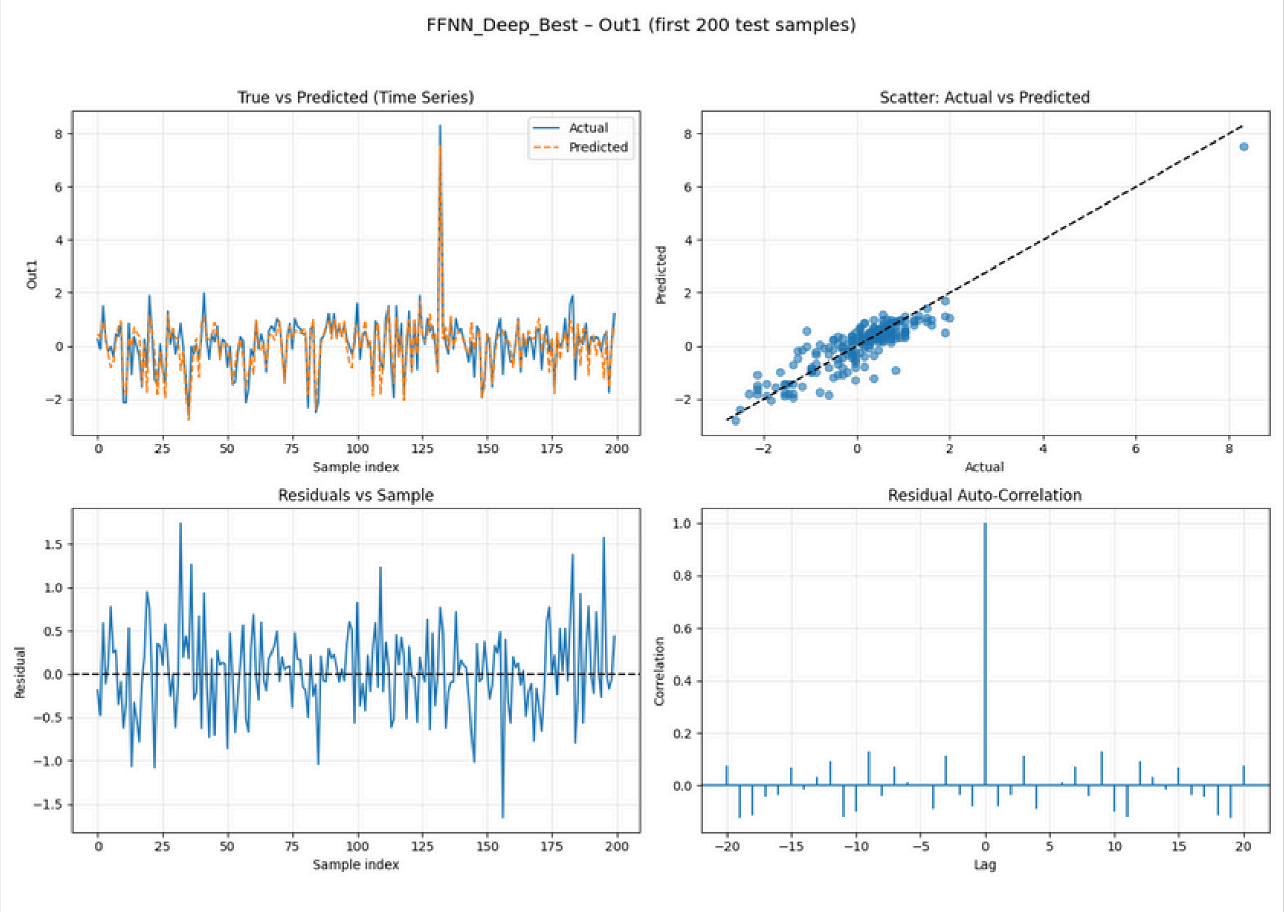
Model Type	Hidden Layers	Learning Rate	L2 Regularization	Best Model Name
Shallow FFNN	1 hidden layer	0.001	0	FFNN_layers1_lr0.001_l20.0
Deep FFNN	3 hidden layers	0.001	0.0001	FFNN_layers3_lr0.001_l20.0001

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Shallow FFNN	0.6093	0.6229	0.6139	1628.22	-4205.28	-4167.05
Deep FFNN	0.7088	0.6849	0.5611	1360.61	-4981.15	-4942.93

ANALYSIS: The deep feedforward neural network performs better than the shallow network for Out1 across all evaluation metrics. While the shallow model achieves a test R^2 of 0.623, the deep model improves it to 0.685, indicating stronger predictive ability. The deep network also shows lower SSE (1360 vs. 1628), reflecting reduced total error. Furthermore, the deep network has significantly lower AIC and BIC values, confirming superior model quality even after accounting for complexity.



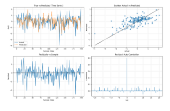
SHALLOW FFNN

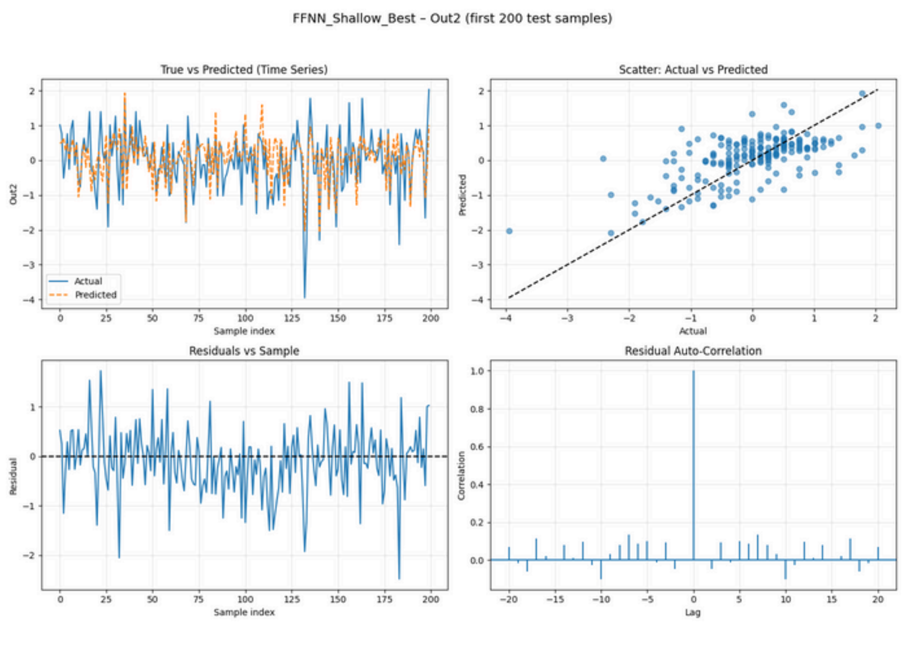


DEEP FFNN

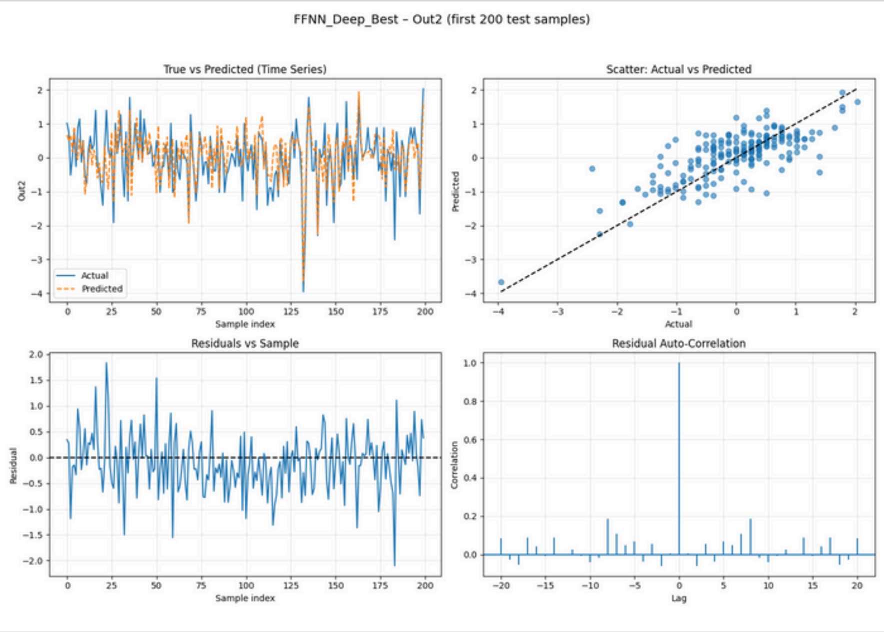
BEST FEEDFORWARD NEURAL NETWORK CONFIGURATIONS FOR OUT2

Model Type	Hidden Layers	Learning Rate	L2 Regularization	Model Name
Shallow FFNN	1 hidden layer	0.001	0	FFNN_layers1_lr0.001_l20.0
Deep FFNN	3 hidden layers	0.001	0.0001	FFNN_layers3_lr0.001_l20.0001

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Shallow FFNN	0.4052	0.3797	0.7712	2570.16	-2232.85	-2194.62
Deep FFNN	0.5251		0.7189	2233.45	-2839.61	-2801.38



SHALLOW FFNN

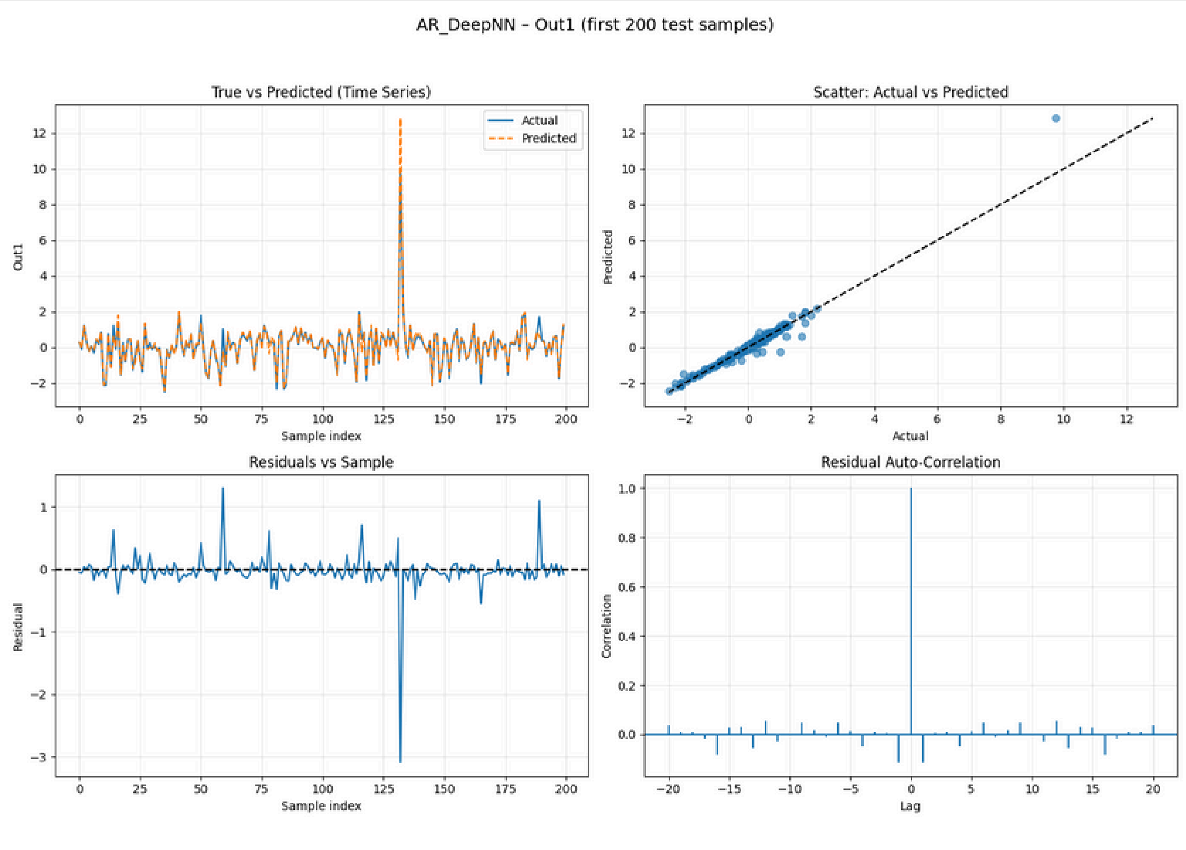


DEEP FFNN

ANALYSIS: It shows better predictive performance for Out2 than the shallow model. The test R^2 improves from 0.380 to 0.461, showing that added depth helps capture nonlinearity. Lower AIC and BIC values for the deep model indicate superior model quality.

AUTOREGRESSIVE DEEP NEURAL NETWORK (OUT1)

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
AR-DeepNN	0.9665	0.9609	0.1971	167.75	-13997.78	-13883.1

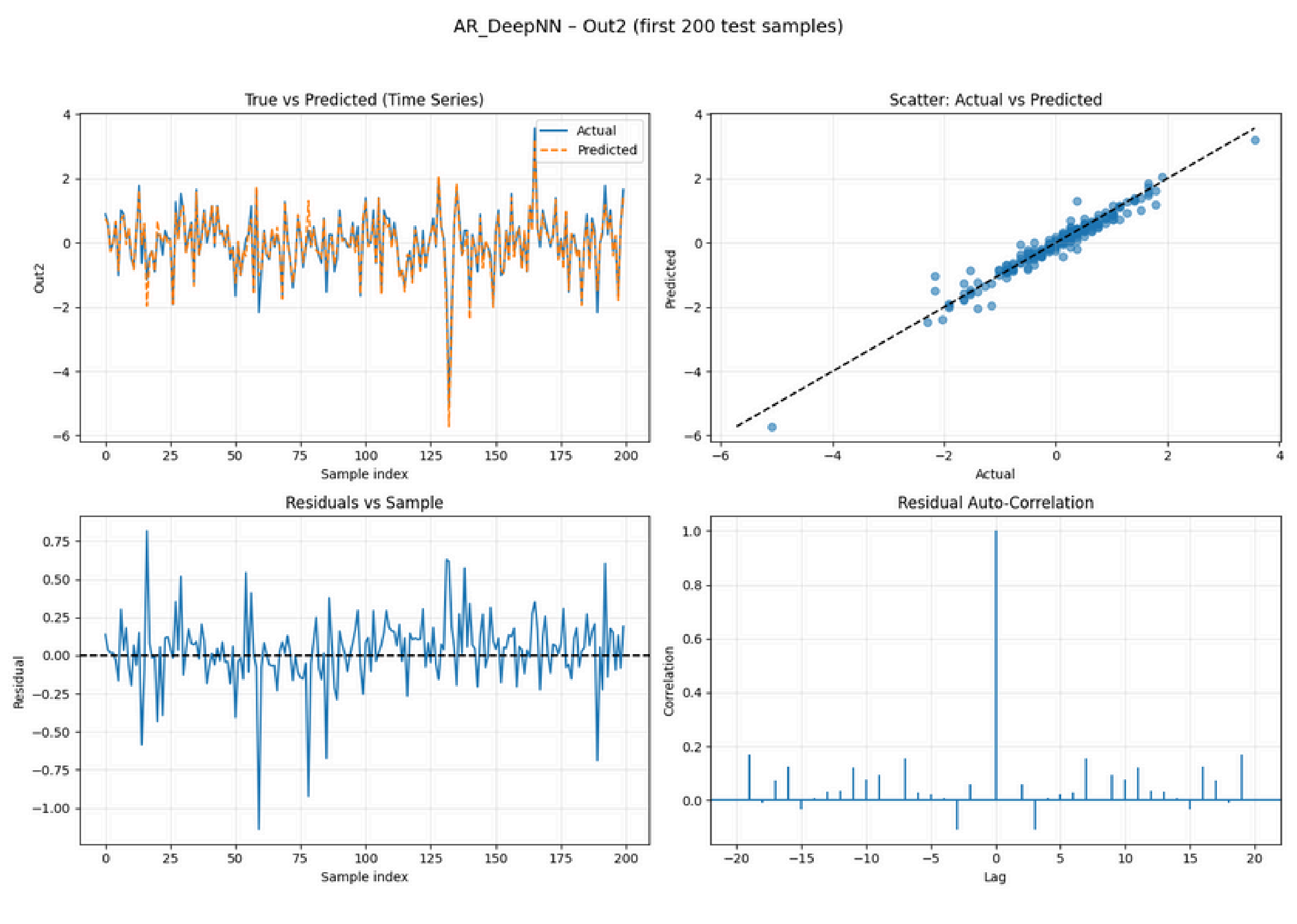


ANALYSIS: The autoregressive deep neural network achieves very high accuracy with a test R^2 of 0.9609, indicating excellent predictive performance. The close agreement between train and test R^2 shows that the model generalizes well and does not overfit.

AUTOREGRESSIVE DEEP NEURAL NETWORK (OUT2)

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
AR-DeepNN	0.9509	0.9412	0.241	250.96	-12257.54	-12142.86

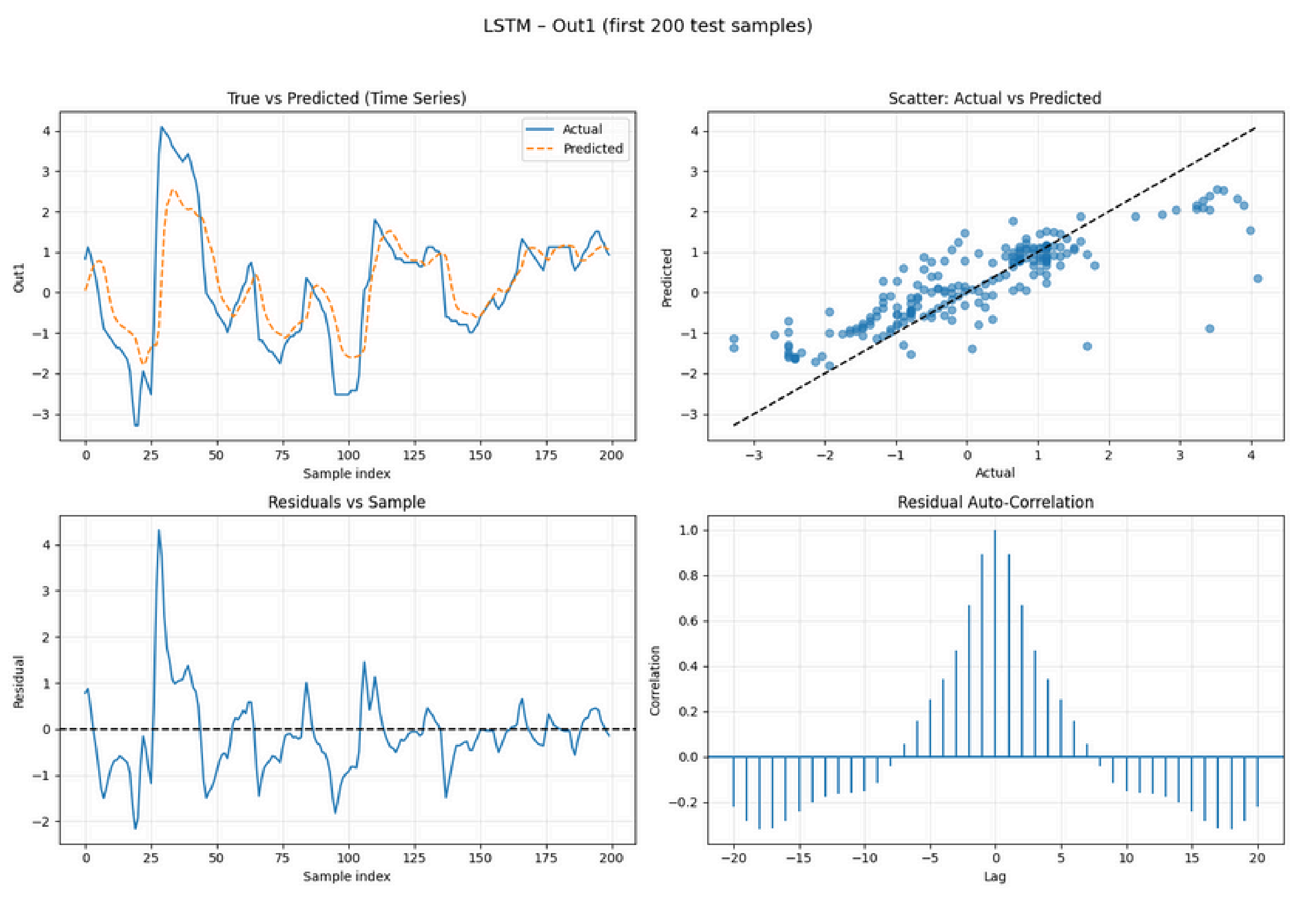
ANALYSIS: It achieves prediction accuracy for Out2 with a test R^2 of 0.941. The close match between training and testing R^2 confirms that the model is not overfitted. The significantly low AIC and BIC values demonstrate superior model quality despite higher complexity.



LSTM MODEL (OUT1)

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
LSTM	0.767	0.5618	0.7682	2548.54	-2206.29	-1976.95

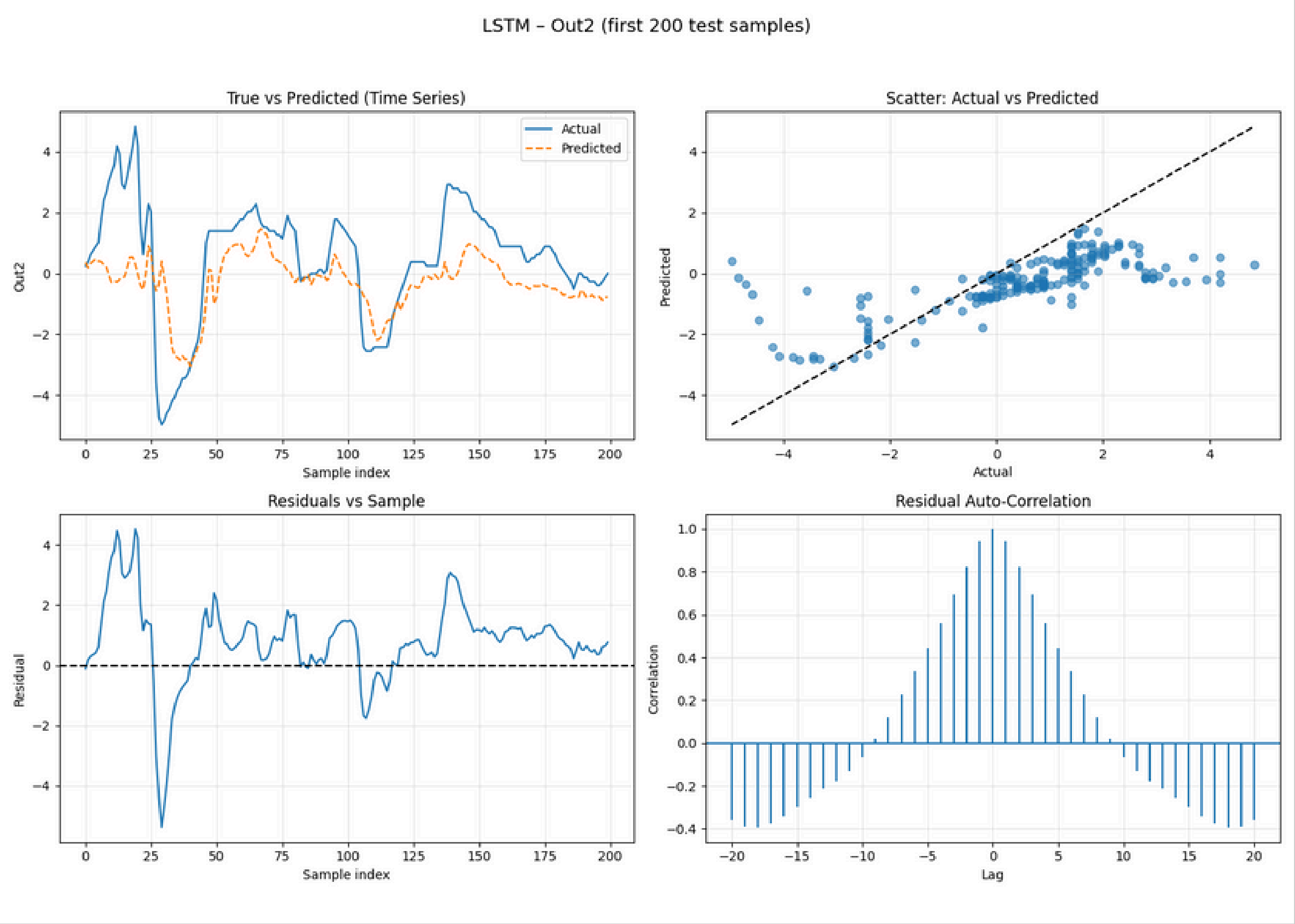
ANALYSIS: The LSTM model shows significantly lower performance compared to the autoregressive deep neural network for Out1 as it uses uses real historical values directly, while LSTM only estimates memory internally.The high SSE and RMSE values suggest large prediction errors. The relatively higher AIC and BIC indicate poorer model fitness.



LSTM MODEL FOR OUT2

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
LSTM	0.7334	0.0285	1.2749	7020.4	2170.15	2399.5

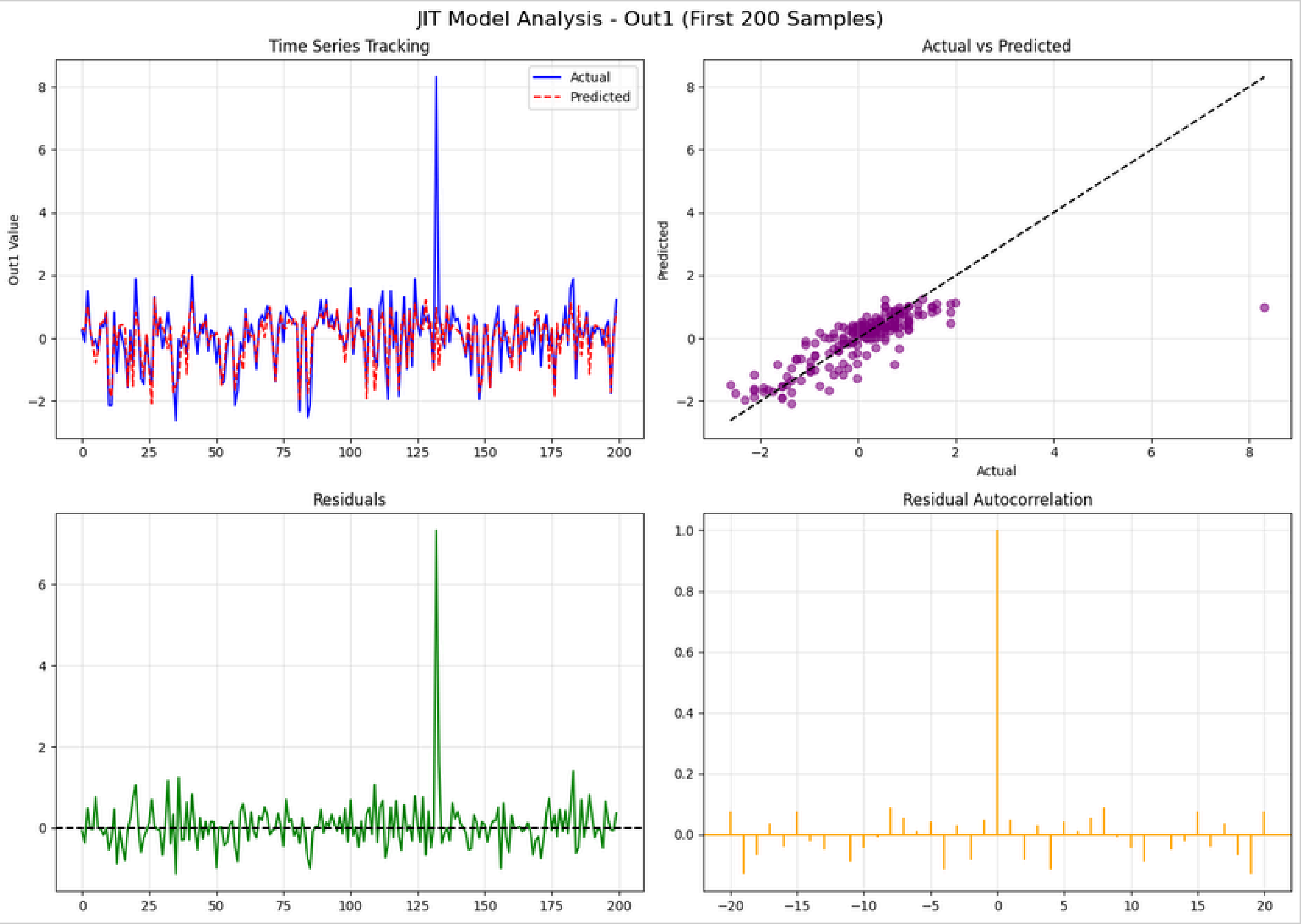
ANALYSIS: The LSTM model performs poorly for Out2, achieving a test R^2 close to zero. The large gap between training and testing R^2 indicates overfitting. High RMSE and SSE values reflect large prediction errors. LSTM expects long-term temporal patterns, but Out2 is more sensitive to current operating conditions. Therefore, the LSTM model is not appropriate for predicting Out2 in this dataset.



JIT-NN MODEL FOR OUT1

Model	Test R^2	RMSE	SSE
JIT Neural Network	0.5911	0.6903	95.31

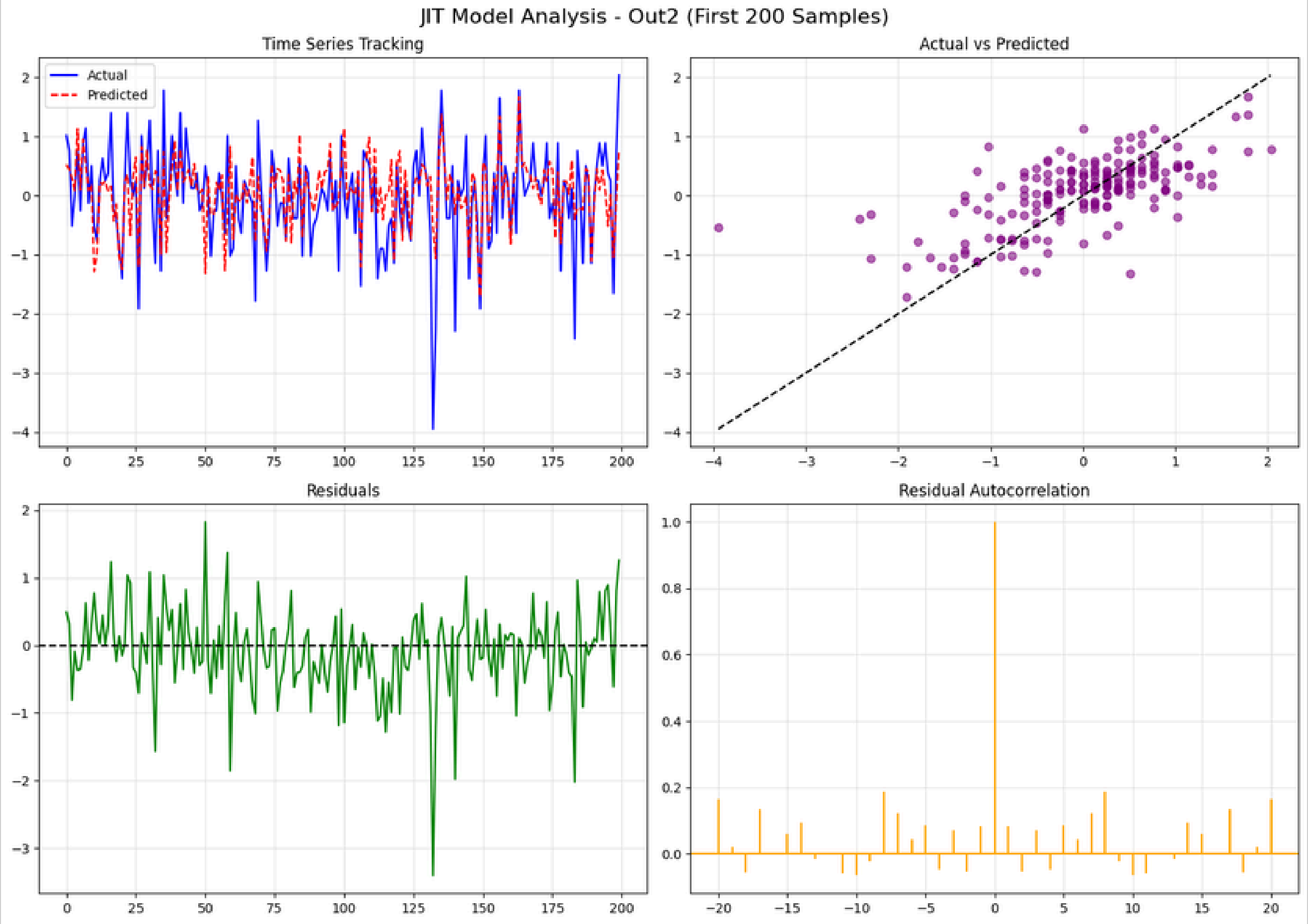
ANALYSIS: The JIT-based neural network demonstrates reasonable prediction accuracy for Out1 with a test R^2 of 0.591. The RMSE indicates moderate prediction errors despite the use of local models. Although SSE is lower than in conventional global ANN models, performance remains inferior to dynamic models such as AR-DeepNN. This shows that local learning alone cannot fully capture SRU process dynamics. Hence, JIT-NN is suitable for approximation but not optimal for Out1 prediction.



JIT-NN MODEL FOR OUT2

Model	Test R^2	RMSE	SSE
JIT Neural Network	0.4154	0.6456	83.36

ANALYSIS: The JIT-based neural network shows moderate predictive capability for Out2. Although it captures some nonlinear behavior, the relatively high RMSE indicates noticeable prediction errors. Hence, the JIT-NN model is less effective than global dynamic models for Out2.



SUMMARY TABLE FOR OUTPUT 1

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Shallow FFNN	0.6093	0.6229	0.6139	1628.22	-4205.28	-4167.05
Deep FFNN	0.7088	0.6849	0.5611	1360.61	-4981.15	-4942.93
AR-DeepNN	**0.9665**	**0.9609**	**0.1971**	**167.75**	**-13997.78**	**-13883.1**
LSTM	0.767	0.5618	0.7682	2548.54	-2206.29	-1976.95
JIT-NN (200 samples)	–	0.5911	0.6903	95.31	–	–

SUMMARY TABLE FOR OUTPUT 2

Model	Train R^2	Test R^2	RMSE	SSE	AIC	BIC
Shallow FFNN	0.4052	0.3797	0.7712	2570.16	-2232.85	-2194.62
Deep FFNN	0.5251	0.461	0.7189	2233.45	-2839.61	-2801.38
AR-DeepNN	**0.9509**	**0.9412**	**0.241**	**250.96**	**-12257.54**	**-12142.86**
LSTM	0.7334	0.0285	1.2749	7020.4	2170.15	2399.5
JIT-NN (200 samples)	–	0.4154	0.6456	83.36	–	–

CONCLUSION

- Out1 exhibits strong nonlinear and dynamic behavior, which could not be fully captured by linear or polynomial regression models.
- Second-order models improved performance, their accuracy remained limited.
- Feedforward neural networks performed better, with deep architectures outperforming shallow ones by capturing greater nonlinearity.
- LSTM did not generalize well due to insufficient temporal dependency and limited data.
- The autoregressive deep neural network (AR-DeepNN) achieved the highest predictive accuracy with minimal error, confirming that Out1 strongly depends on both current states and historical information. Therefore, AR-DeepNN is selected as the optimal soft sensor model for Output 1.

- **Out2** proved significantly more challenging for static and sequence-based models due to weak linearity and irregular dynamics.
- All regression and PCA-based models performed poorly, indicating that Out2 cannot be explained by instantaneous relationships alone.
- Deep feedforward networks showed improvement, accuracy remained moderate.
- LSTM model failed to generalize due to lack of consistent sequence patterns and insufficient data for sequence learning.
- In contrast, the autoregressive deep neural network captured both nonlinear and delayed effects effectively, leading to very high test accuracy.
- Hence, AR-DeepNN is identified as the most reliable predictive model for Output 2.

SUGGESTED IMPROVEMENTS

KEY LEARNINGS:

- Practical understanding of how machine learning can replace physical sensors through soft sensors.
- Learned that model choice matters more than complexity – the best model is not always the most advanced one.
- Discovered the importance of data preprocessing such as scaling and feature selection.
- Understood how nonlinear models outperform linear ones in real industrial systems.
- Learned that time-dependence must be tested, not assumed.
- Importance of performance metrics (R^2, RMSE, SSE, AIC, BIC) in selecting the best model.

IMPROVEMENTS:

- Increase the number of lab sessions along with theory classes.
- Provide hands-on implementation for every major concept explained in lectures.
- Include practical evaluation through coding assignments and mini-projects.
- Use more real-world industrial datasets for better understanding.
- Conduct model comparison as part of labs to strengthen learning.