Optimizing Latent Dimension for Long-Horizon Forecasting in Nuclear Power Plants Using Latent Prediction Models

Donghee Jung[a], Junyong Bae[b], and Seung Jun Lee[c]

[a] Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, joy00@unist.ac.kr
[b] Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, junyong0513@unist.ac.kr
[c] Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea, sjlee420@unist.ac.kr

Abstract: Long-horizon forecasting in nuclear power plants is essential for operator support under abnormal and accident conditions, where timely understanding of future parameter evolution can improve decision-making and help maintain safety margins. In such situations, operator support systems should not only interpret the current plant state but also anticipate the future trends of multiple key variables over extended time horizons. However, as the forecasting horizon becomes longer and the number of variables increases, direct forecasting models often suffer from degraded performance due to high-dimensional correlations and nonlinear dynamic behavior. To address this challenge, this study investigates a latent prediction framework for long-horizon forecasting and focuses on optimizing the latent dimension, which is a key design variable in latent-space modeling.

A pilot study on nuclear power plant emergency scenario forecasting showed that a two-stage latent-space approach, in which an autoencoder compresses future parameter trends into latent vectors and a predictor estimates those latent vectors before reconstruction through a decoder, can outperform direct forecasting models in both accuracy and efficiency for long-horizon prediction. Building on this motivation, the present study treats the latent dimension not as a fixed empirical parameter but as an optimization target. Multiple latent dimension candidates were evaluated by jointly considering reconstruction quality and forecasting performance.

The results indicate that an excessively small latent dimension fails to preserve important dynamic characteristics and inter-variable relationships, leading to reduced long-horizon prediction accuracy. In contrast, an excessively large latent dimension introduces unnecessary degrees of freedom, which can weaken generalization performance and reduce learning efficiency. The study further suggests that the optimal latent dimension depends on the complexity of the forecasting problem and tends to increase as the number of target variables grows. These findings highlight that latent-space capacity should be adaptively determined according to the dimensionality of the plant variables and the complexity of the prediction task.

This work contributes by emphasizing latent dimension optimization as a critical step in developing latent prediction models for nuclear power plant applications. The proposed perspective can support future applications such as trip margin prediction, mitigation action assessment, and advanced operator support for long-horizon, high-dimensional forecasting problems.

Keywords: latent dimension optimization; long-horizon forecasting; nuclear power plant; operator support; LSTM autoencoder; latent prediction model

1. INTRODUCTION

Accurate forecasting of nuclear power plant parameters is important for operator support during abnormal and accident conditions. In such situations, operators need to understand not only the current plant state but also how major plant variables are expected to evolve over the next several minutes or hours. Forecasts of reactor power, loop temperatures, pressurizer pressure, steam-generator levels, containment variables, and related parameters can support situation awareness and help operators assess whether sufficient safety margins are being maintained.

Data-driven forecasting models have been investigated as fast surrogate models for predicting plant parameter trends. Previous studies have shown that recurrent neural networks, long short-term memory models, and related machine learning approaches can be used to forecast nuclear power plant responses from short observation histories or scenario information [1-4]. However, many forecasting formulations directly predict future values in the original observation space. This direct formulation becomes increasingly challenging when both the forecast horizon and the number of target variables increase. For a forecast horizon H and V target variables, the model output has size H × V. For example, a 1500-step forecast of 25 plant parameters requires the model to predict 37,500 future values from a short input history.

Latent-space forecasting is a promising way to reduce this output-space burden. Instead of directly predicting the full future trajectory, a latent prediction model first represents future trajectories using compact latent vectors. A separate predictor then estimates these latent vectors from short observation histories, and a decoder reconstructs the full future trajectory. This approach changes the prediction target from a high-dimensional trajectory to a lower-dimensional trajectory coordinate.

However, the latent dimension is a critical design variable. If the latent dimension is too small, important transient characteristics and inter-variable relationships may be lost. If the latent dimension is unnecessarily large, the latent representation may include redundant degrees of freedom and reduce training efficiency. Therefore, latent dimension should be selected according to the complexity of the forecasting task rather than fixed empirically.

This paper focuses on latent-dimension selection for long-horizon multivariate nuclear power plant parameter forecasting. The main objective is not to propose a new neural network architecture, but to examine how the latent dimension affects forecasting performance under different output-dimensionality conditions. The contributions of this paper are as follows. First, long-horizon plant parameter forecasting is formulated as a latent prediction problem. Second, the effect of latent dimension is evaluated in a 25-channel, 1500-step forecasting task. Third, a reduced 10-channel task is analyzed to examine how output-variable count affects the required latent capacity.

2. LATENT PREDICTION FRAMEWORK

This study considers multivariate forecasting from a short observation history. Let X denote the observed input sequence and Y denote the future target trajectory:

$$X \in \mathbb{R}^{L \times N_{in}}, \; Y \in \mathbb{R}^{H \times V} \tag{1}$$

where L is the input length, N_{in} is the number of input variables, H is the forecast horizon, and V is the number of target variables.

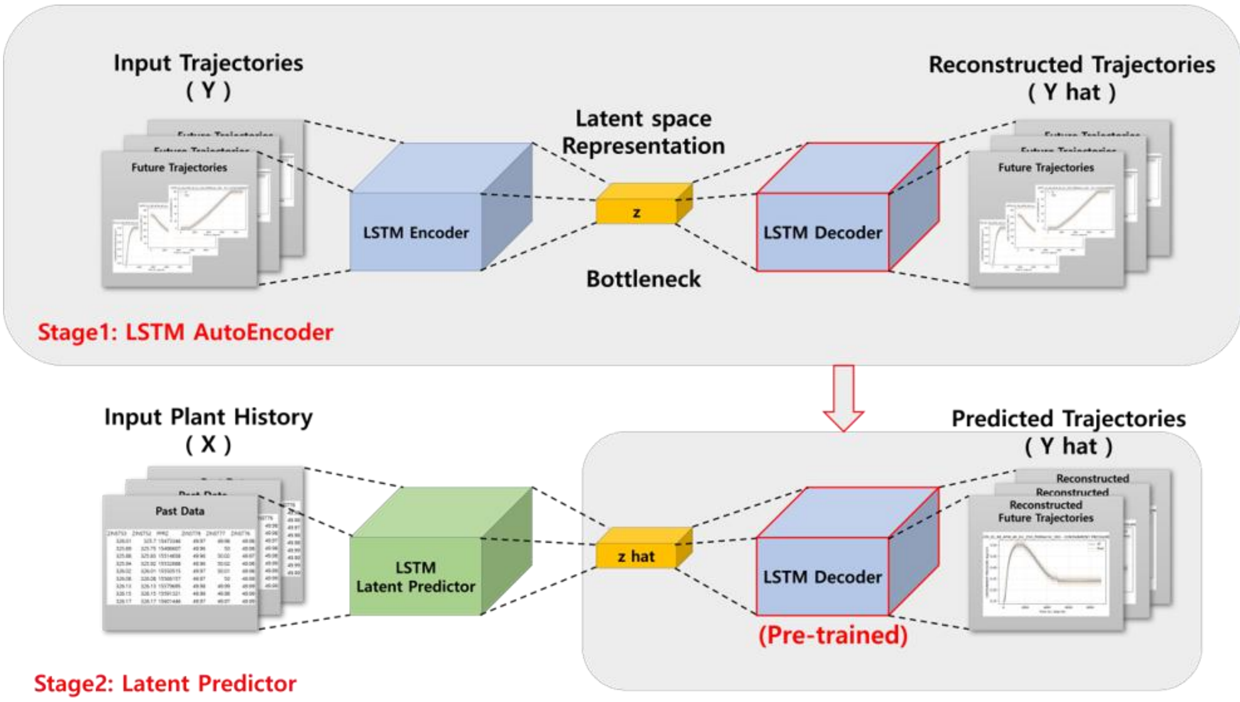
In direct observation-space forecasting, a neural network learns a direct mapping from X to Y:

$$\hat{Y} = f_{direct}(X) \tag{2}$$

This formulation is simple, but the model must directly estimate the full H × V future trajectory. As H and V increase, the model output becomes large and contains complex temporal and inter-variable dependencies.

The latent prediction framework separates trajectory representation learning from short-history prediction. Figure 1 shows the two-stage framework used in this study. In the first stage, an LSTM autoencoder compresses each future trajectory Y into a latent vector z:

Figure 1: Latent prediction framework.



$$z = Encoder(Y) \tag{3}$$

$$\tilde{Y} = Decoder(z) \tag{4}$$

where $z \in \mathbb{R}^{D_z}$, and D_z is the latent dimension. The autoencoder is trained to reconstruct future trajectories from their latent representations.

In the second stage, a predictor maps the short observation history X to the latent vector:

$$\hat{z} = Predictor(X) \tag{5}$$

The final future trajectory is obtained by passing the predicted latent vector through the pre-trained decoder:

$$\hat{Y} = Decoder(\hat{z}) \tag{6}$$

In this structure, the predictor does not directly regress all H × V output values. Instead, it predicts a compact latent coordinate that represents the overall future trajectory. The decoder then reconstructs the multivariate trajectory from that coordinate. Therefore, D_z controls the capacity of the latent trajectory representation and directly affects the ability of the model to preserve transient shapes, delayed responses, long-term trends, and inter-variable relationships.

The central question of this paper is how D_z should be selected for long-horizon multivariate forecasting. A very small D_z may create an under-capacity bottleneck, whereas a very large D_z may provide limited additional benefit. Therefore, the goal is to identify a practical operating range of D_z for each forecasting task.

3. EXPERIMENTAL SETUP

The experiments were conducted using compact nuclear simulator data for operator-support-oriented forecasting. The raw simulator data were recorded at 1 s resolution and sampled at 6 s intervals for the forecasting tasks. The main forecast horizon was H = 1500, corresponding to approximately 9000 s, or 2.5 h, of future prediction. Each input sample used five observed time steps constructed from 109 plant variables, including target plant parameters and additional plant-state variables.

Two output configurations were considered. The first was a 25-channel task including representative plant parameters such as reactor power, neutron range indicators, core and loop temperatures, pressurizer pressure and level, steam-generator levels and pressures, feedwater flows, and containment variables.

The second was a reduced 10-channel task selected from the 25-channel set to retain plant-wide subsystem coverage while reducing repeated loop-wise variables.

The latent forecasting model used an LSTM autoencoder and an LSTM-based history-to-latent predictor. The autoencoder consisted of an LSTM encoder and decoder. After autoencoder training, the decoder was fixed and the predictor was trained to estimate the latent vector from the short observation history. The latent dimension D_z was varied to evaluate its influence on long-horizon forecasting performance.

The main evaluation metrics were mean absolute error (MAE), root mean squared error (RMSE), and dynamic time warping (DTW). In addition, a tolerance-based success rate was used as a supplementary indicator. The success rate was computed as the proportion of trajectory-variable cases in which no more than 10% of predicted points fell outside a +/-10% range-based tolerance band. RMSE was treated as the primary error metric because it is sensitive to large deviations in long-horizon forecasts.

Table 1: Summary of latent-dimension sweep tasks.

Task	Forecast Horizon	Output Channels	Input Shape	Latent Dimensions
25-channel z sweep	1500	25	5 × 109	4, 8, 16, 24, 32, 48, 64
Reduced-output z sweep	1500	10	5 × 109	4, 8, 16, 24, 32, 48, 64

4. RESULTS AND DISCUSSION

4.1. Latent-Dimension Effect in the 25-Channel Task

Table 2 summarizes the latent-dimension sweep results for the 25-channel, 1500-step forecasting task. In this conference paper, the analysis focuses on compact and intermediate latent dimensions from D_z = 4 to D_z = 64, because this range is most relevant to practical latent-space model configuration.

Table 2: Latent-dimension sweep results for the 25-channel task.

D_z	MAE	RMSE	DTW	Tolerance-Based Success Rate (%)
4	0.02630	0.07174	0.003002	89.26
8	0.01228	0.03054	0.001188	97.26
16	0.01093	0.02779	0.001109	98.06
24	0.01090	0.02813	0.001058	98.51
32	0.01035	0.02520	0.001065	98.40
48	0.01116	0.02659	0.001124	97.94
64	0.01029	0.02542	0.001065	98.29

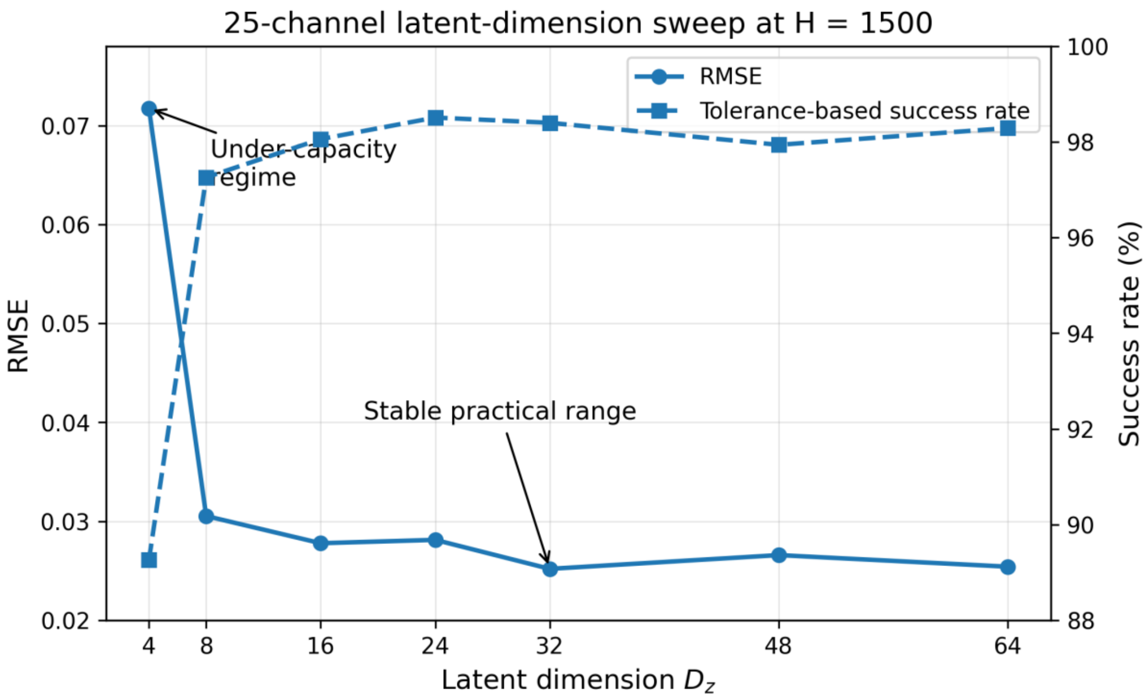
The results show a clear under-capacity regime at D_z = 4. In this case, RMSE increased to 0.07174 and the tolerance-based success rate decreased to 89.26%. This indicates that a four-dimensional latent vector was not sufficient to represent the diversity of 25-channel, 1500-step future trajectories. Important transient characteristics and inter-variable relationships were likely compressed too strongly.

When D_z increased from 4 to 8, the prediction performance improved sharply. RMSE decreased from 0.07174 to 0.03054, and the tolerance-based success rate increased from 89.26% to 97.26%. This suggests that the latent representation exited the under-capacity regime once the latent space had sufficient dimensions to encode the dominant trajectory patterns.

For D_z values from 16 to 64, the performance remained relatively stable. The lowest RMSE was obtained at D_z = 32, while D_z = 64 produced the lowest MAE. The highest tolerance-based success rate was observed at D_z = 24. Therefore, the results do not indicate a single sharply optimal latent dimension. Instead, they suggest that intermediate latent dimensions form a practical operating range for the 25-channel long-horizon forecasting task.

Figure 2 summarizes the 25-channel sweep results. The same axis ranges are used in Figure 2 and Figure 3 to allow direct comparison between the 25-channel and reduced 10-channel tasks. The sharp difference between $D_z = 4$ and the intermediate latent dimensions supports the interpretation that overly aggressive compression can distort long-horizon multivariate trajectory representations.

Figure 2: Quantitative summary of the 25-channel sweep.



4.2. Latent-Dimension Effect in the Reduced 10-Channel Task

Table 3 shows the latent-dimension sweep results for the reduced 10-channel task. This experiment was conducted to examine whether reducing the number of output variables makes very small latent dimensions sufficient.

Table 3: Latent-dimension sweep results for the reduced 10-channel task.

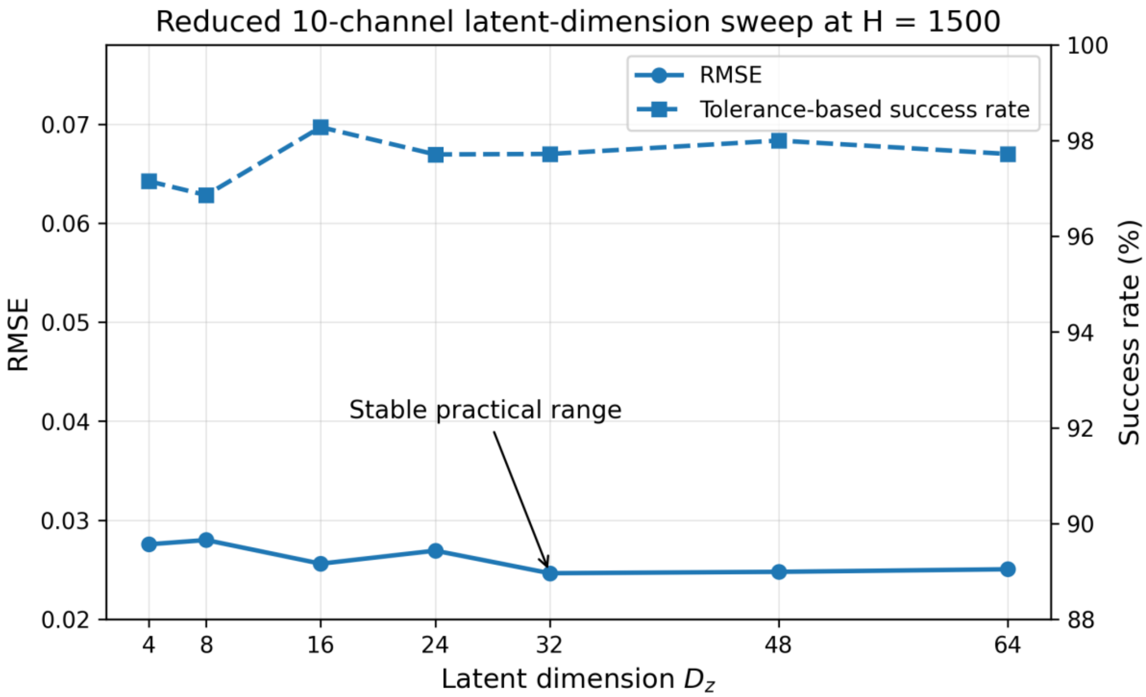
D_z	MAE	RMSE	DTW	Tolerance-Based Success Rate (%)
4	0.01265	0.02757	0.001339	97.15
8	0.01326	0.02800	0.001371	96.86
16	0.01175	0.02561	0.001273	98.28
24	0.01166	0.02693	0.001311	97.71
32	0.01107	0.02465	0.001287	97.72
48	0.01142	0.02479	0.001285	98.00
64	0.01083	0.02505	0.001187	97.72

Compared with the 25-channel task, the reduced 10-channel task showed much weaker degradation at small latent dimensions. The RMSE at $D_z = 4$ was 0.02757, which is far lower than the corresponding 25-channel value of 0.07174. This indicates that reducing the number of output variables reduces the burden on the latent representation.

However, the smallest latent dimension was not clearly the best choice. The lowest RMSE was obtained at $D_z = 32$, while the lowest MAE and DTW were obtained at $D_z = 64$. The highest success rate was obtained at $D_z = 16$. Therefore, even in the reduced-output case, intermediate latent dimensions provided stable overall performance.

Figure 3 summarizes the reduced 10-channel sweep results using the same axis ranges as Figure 2. Unlike the 25-channel task, the reduced-output task does not show a distinct under-capacity regime at the smallest latent dimensions. This supports the interpretation that output-variable count affects the severity of the bottleneck effect. However, the reduced task still benefits from sufficient latent capacity, showing that the required D_z is not determined by the number of output variables alone.

Figure 3: Quantitative summary of the reduced 10-channel sweep.



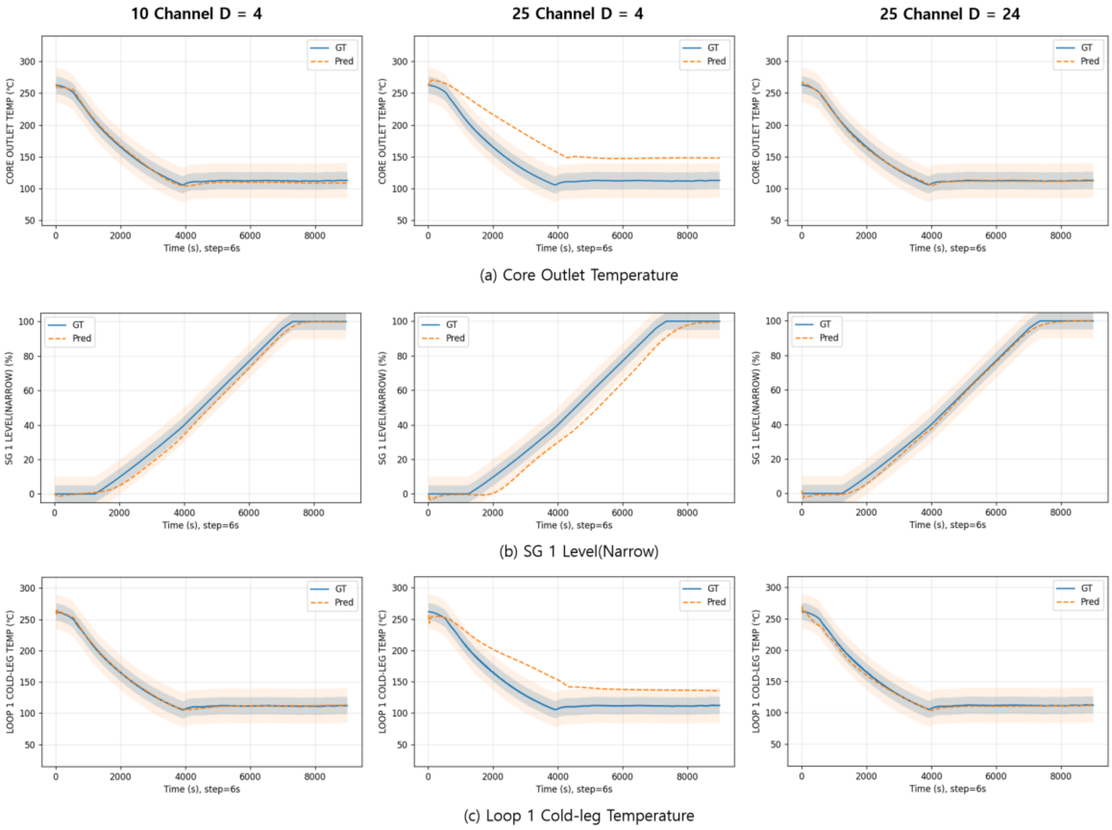
4.3. Trajectory-Level Interpretation

The aggregate results indicate that the effect of latent dimension depends on both output dimensionality and trajectory complexity. To illustrate this point, Figure 4 compares representative trajectories for selected variables in an SGTR scenario. The comparison includes three settings: a reduced 10-channel model with $D_z = 4$, a 25-channel model with $D_z = 4$, and a 25-channel model with $D_z = 24$.

The trajectory comparison shows that some slowly varying or nearly constant variables can be reconstructed reasonably well even with a small latent dimension. However, variables with nonlinear cooling, pressure rise and decay, delayed level increase, or accident-dependent recovery behavior require more latent capacity. In the 25-channel $D_z = 4$ case, several complex transients are over-smoothed or biased, indicating that the latent representation is not sufficiently expressive. Increasing the latent dimension to $D_z = 24$ recovers these trajectory structures more effectively.

These qualitative results are consistent with the quantitative trends in Tables 2 and 3. The smallest bottleneck fails primarily when task complexity and trajectory complexity jointly exceed the representational capacity of the latent space. Reducing the number of output variables mitigates this problem, but the latent representation still benefits from an intermediate capacity range.

Figure 4: Representative trajectory comparison.



4.4. Practical Implications

The results support three practical observations for long-horizon nuclear plant parameter forecasting.

First, latent dimension should be treated as a model design variable rather than a fixed empirical hyperparameter. In the 25-channel task, D_z = 4 was clearly insufficient, while intermediate values produced stable performance. Therefore, selecting the smallest possible latent dimension is not always desirable.

Second, the appropriate latent dimension depends on the complexity of the forecasting task. When the number of output variables was reduced from 25 to 10, the low-dimensional degradation became much less severe. This means that output dimensionality is one important factor in latent-dimension selection. However, the reduced task still benefited from intermediate latent dimensions, showing that trajectory complexity must also be considered.

Third, practical selection should focus on a reliable operating range rather than a single best value. In the 25-channel task, D_z values around 24 to 64 provided strong overall performance. In the reduced 10-channel task, D_z values around 16 to 64 were generally stable. Since different metrics favored slightly different latent dimensions, selecting D_z based on one metric alone may be misleading. A more robust strategy is to evaluate reconstruction quality, forecasting error, trajectory-shape similarity, and tolerance-based success rate together.

For future operator-support systems, these findings suggest that latent-space forecasting can provide a compact and efficient representation of long future trajectories, but only if the latent space is sufficiently expressive. A latent dimension that is too small may produce over-smoothed or biased trajectories, which could be problematic in safety-related forecasting applications. Therefore, latent-dimension selection should be included as an explicit step in the model development process.

5. CONCLUSION

This paper examined latent-dimension selection for long-horizon multivariate forecasting of nuclear power plant parameters. The forecasting problem was formulated as a latent prediction problem, where future plant trajectories are compressed into latent vectors and reconstructed through a decoder. The main focus was the effect of latent dimension D_z on prediction performance.

The 25-channel, 1500-step forecasting task showed a clear under-capacity regime at $D_z = 4$. Increasing the latent dimension sharply improved RMSE, MAE, DTW, and tolerance-based success rate. Intermediate latent dimensions, especially around $D_z = 24$ to 64, provided stable practical performance. The reduced 10-channel task showed that decreasing the number of output variables mitigates low-dimensional degradation, but does not remove the need for sufficient latent capacity.

Overall, the results indicate that latent dimension should be adaptively selected according to the forecast horizon, output-variable count, and trajectory complexity. For long-horizon operator-support forecasting, latent-space models should not be configured only for maximum compression. Instead, they should use enough latent capacity to preserve important dynamic characteristics and inter-variable relationships. Future work will examine repeated training runs, broader accident scenarios, uncertainty-aware latent forecasting, and adaptive latent representations for variable-specific plant behavior.

ACKNOWLEDGEMENTS

This research was supported by the National Research Council of Science & Technology(NST) grant by the Korea government (MSIT) (No. GTL24031-000)

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2022-00143695)

REFERENCES

[1] J. Bae, G. Kim, and S. J. Lee. "Real-time prediction of nuclear power plant parameter trends following operator actions," Expert Systems with Applications, volume 186, 115848, (2021).

[2] J. Bae, K. S. Kim, and S. J. Lee. "Two-stage deep learning model for nuclear power plant parameter trend prediction with epistemic uncertainty quantification," Applied Soft Computing, volume 186, 114233, (2026).

[3] M. I. Radaideh, C. Pigg, T. Kozlowski, Y. Deng, and A. Qu. "Neural-based time series forecasting of loss of coolant accidents in nuclear power plants," Expert Systems with Applications, volume 160, 113699, (2020).

[4] H. Kim and J. Kim. "Long-term prediction of safety parameters with uncertainty estimation in emergency situations at nuclear power plants," Nuclear Engineering and Technology, volume 55, pp. 1630-1643, (2023).

[5] S. Ben Taieb, G. Bontempi, A. F. Atiya, and A. Sorjamaa. "A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition," Expert Systems with Applications, volume 39, pp. 7067-7083, (2012).

[6] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse. "Methodology for long-term prediction of time series," Neurocomputing, volume 70, pp. 2861-2869, (2007).

[7] N. Nguyen and B. Quanz. "Temporal latent auto-encoder: A method for probabilistic multivariate time series forecasting," arXiv, arXiv:2101.10460, (2021).

[8] N. R. Franco, D. Fraulin, A. Manzoni, and P. Zunino. "On the latent dimension of deep autoencoders for reduced order modeling of PDEs parametrized by random fields," arXiv, arXiv:2310.12095, (2023).