# Diffusion-Centric Diffusion Models for Time Series Forecasting

## DMQA Open Seminar

2026.03.27

Data Mining & Quality Analytics Lab.

정구진

# 발표자 소개

## 정구진

- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- Ph.D. Student (2023.03 ~ Present)

## Research Interest

- Multivariate Time Series Modeling
- Generative models

## Contact

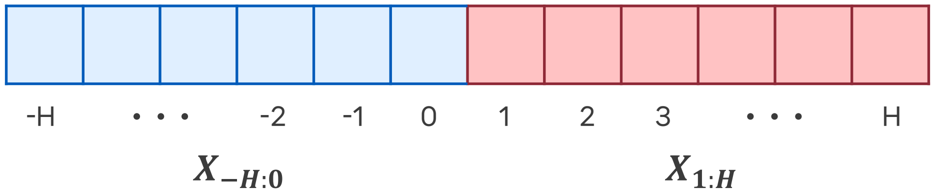
- kujhin@korea.ac.kr

# TABLE OF CONTENTS

# 01 Diffusion Models for Time Series Forecasting (TSF)

## Diffusion Models for TSF

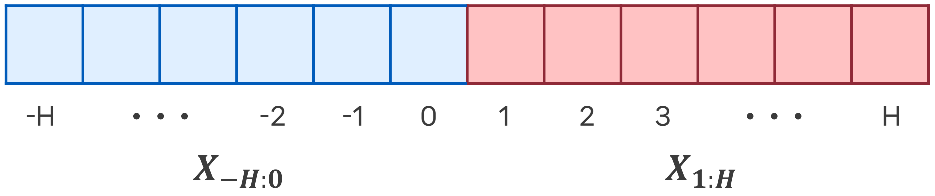
- 시계열 예측을 위한 diffusion 모델은 보통 conditional generative model로 정의
- 과거 관측값 및 추가적인 context 정보를 조건으로 하여 미래 sequence를 생성하는 것이 목적

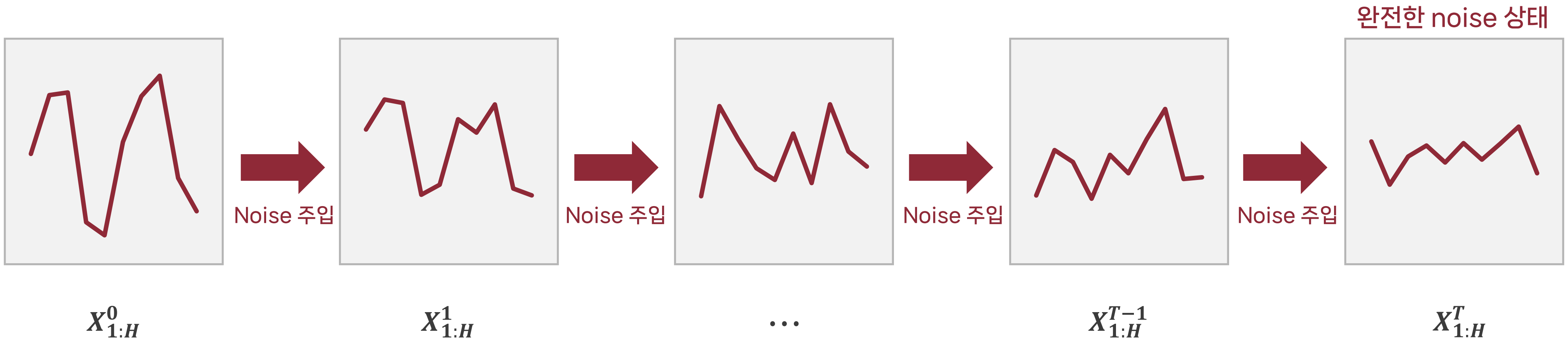


# 01 Diffusion Models for Time Series Forecasting (TSF)

## Diffusion Models for TSF

- 시계열 예측을 위한 diffusion 모델은 보통 conditional generative model로 정의
- 과거 관측값 및 추가적인 context 정보를 조건으로 하여 미래 sequence를 생성하는 것이 목적

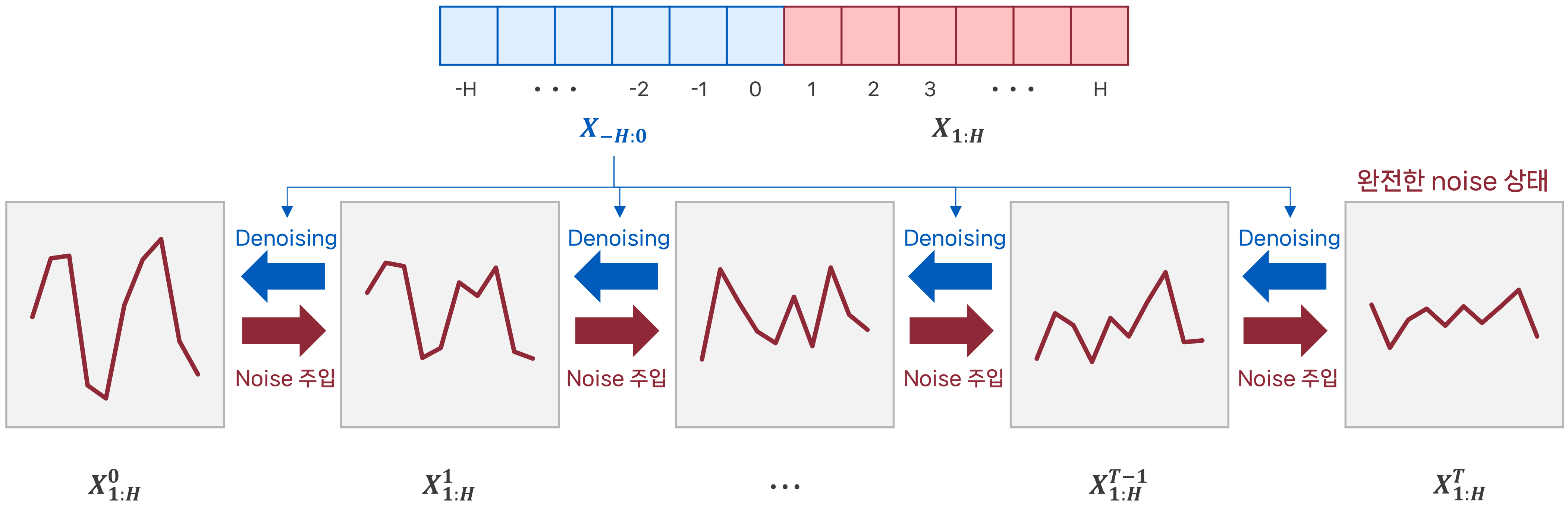


# 01 Diffusion Models for Time Series Forecasting (TSF)

## Diffusion Models for TSF

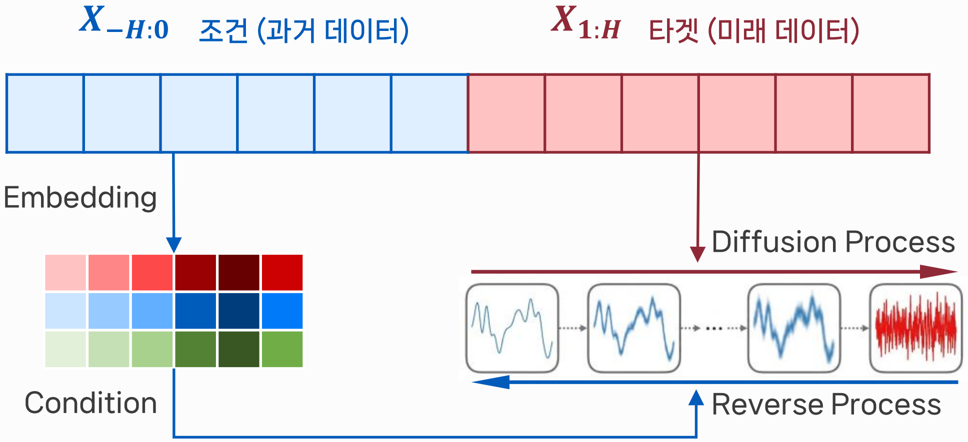
- 시계열 예측을 위한 diffusion 모델은 보통 conditional generative model로 정의
- 과거 관측값 및 추가적인 context 정보를 조건으로 하여 미래 sequence를 생성하는 것이 목적



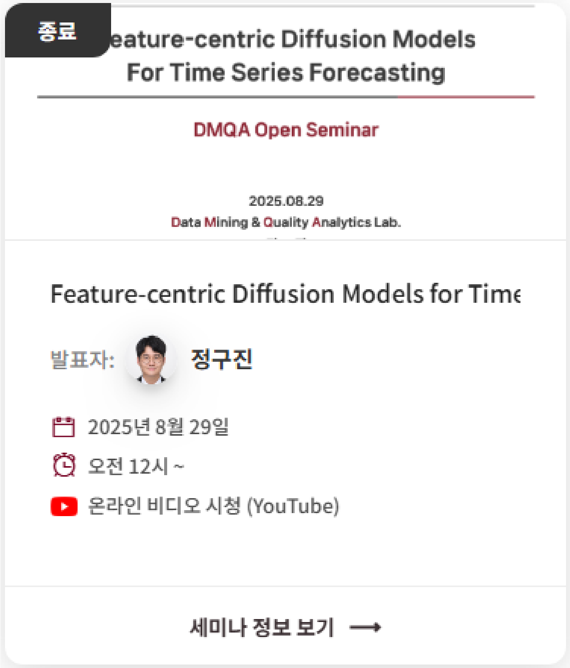
# 01 Two Strategies for Condition Integration

## Feature-centric

Diffusion-centric




- Diffusion process는 그대로 유지하면서 조건 정보를 효과적으로 추출하기 위해 노력
- Recurrent, Convolution, Attention, MLP 등 다양한 방식으로 시계열의 feature를 추출

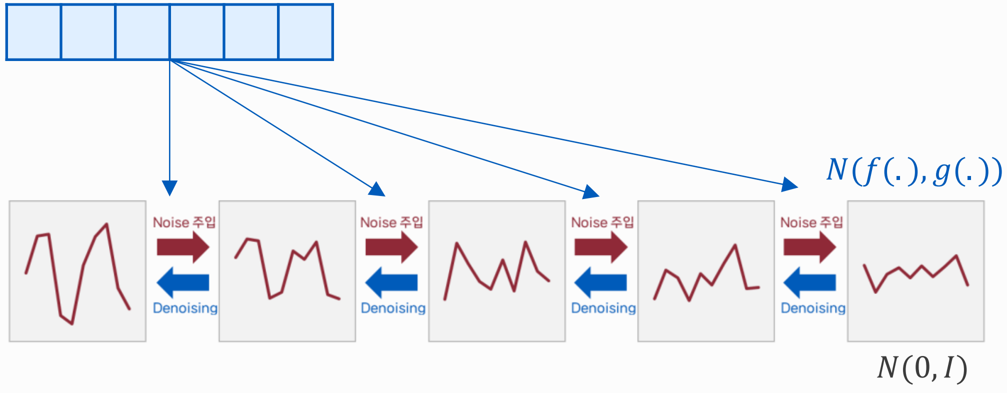


# 01 Two Strategies for Condition Integration

Feature-centric

**Diffusion-centric**

### Enhanced Process




### Reformulated Process

## REDI: Recurrent Diffusion Model for Probabilistic Time Series Forecasting (CIKM, 2024)

# REDI: Recurrent Diffusion Model for Probabilistic Time Series Forecasting

Shiyang Zhou
Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University
Shanghai, China
22210240079@m.fudan.edu.cn

Zehao Gu
Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University
Shanghai, China
guzh22@m.fudan.edu.cn

Yun Xiong*
Shanghai Key Laboratory of Data Science, School of Computer Science, Fudan University
Shanghai, China
yunx@fudan.edu.cn

Yang Luo
MoE Key Lab of Artificial Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University
Shanghai, China
floating-dream@sjtu.edu.cn

Qiang Wang
Shanghai Center for Meteorological Disaster Prevention Technology
Shanghai, China
18918206933@qq.com

Xiaofeng Gao*
MoE Key Lab of Artificial Intelligence, Department of Computer Science and Engineering, Shanghai Jiao Tong University
Shanghai, China
gao-xf@cs.sjtu.edu.cn

### Abstract

Time series forecasting (TSF) consists of point prediction and probabilistic forecasting. Unlike point forecasting which predicts an expected value of a future target, probabilistic time series forecasting models the uncertainty in data by predicting the distribution of future values, which enhances decision-making flexibility and improves risk management. Traditional probabilistic forecasting

across 12 baselines, which strongly demonstrates the effectiveness of REDI.

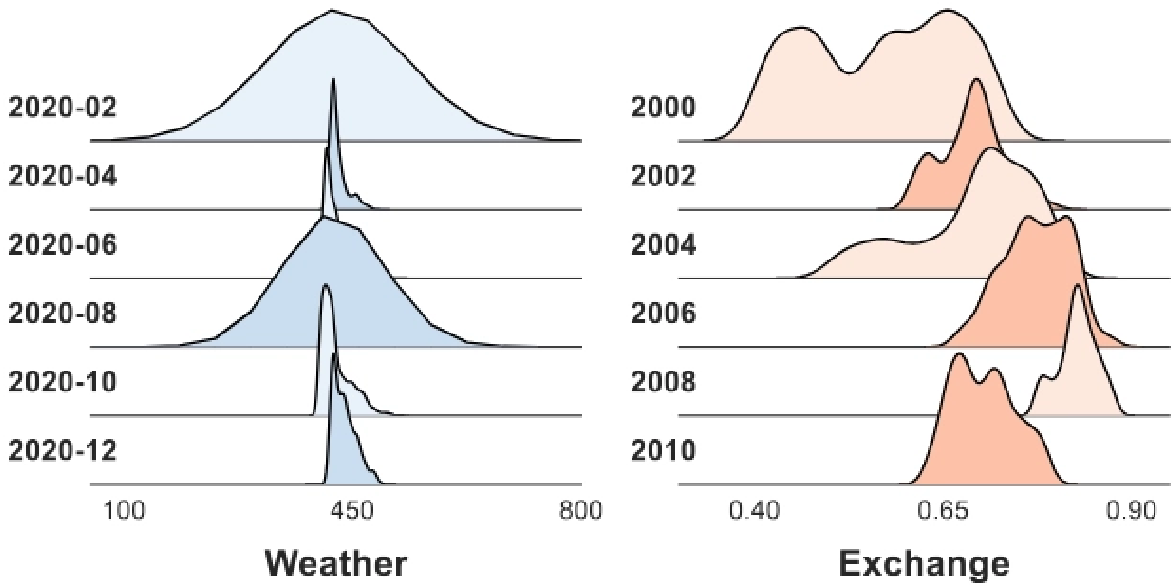
### CCS Concepts

• **Information systems → Data mining;** • **Computing methodologies → Machine learning.**

# 02 연구 배경

## 연구 배경1. Distribution Drift

- 시계열 데이터는 데이터간 분포의 불일치가 발생
- Diffusion 기반의 기존 방법들은 예측대상이 되는 타겟에만 점진적으로 노이즈를 주입하여 이러한 시계열 패턴을 포착하지 못함

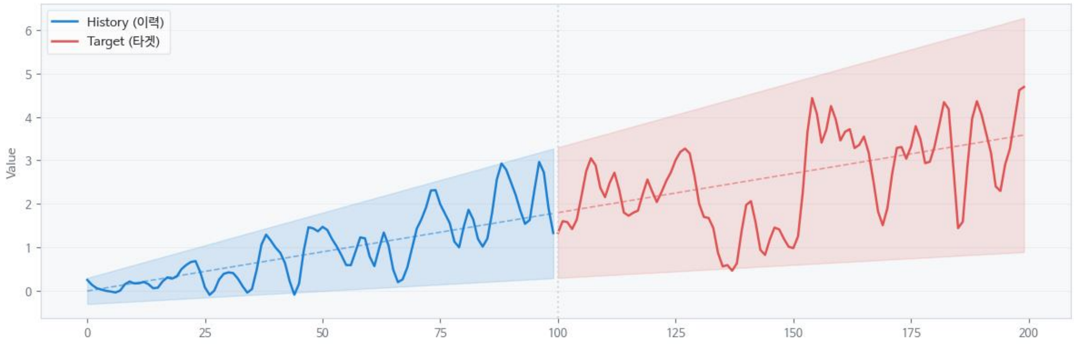


# 02 연구 배경

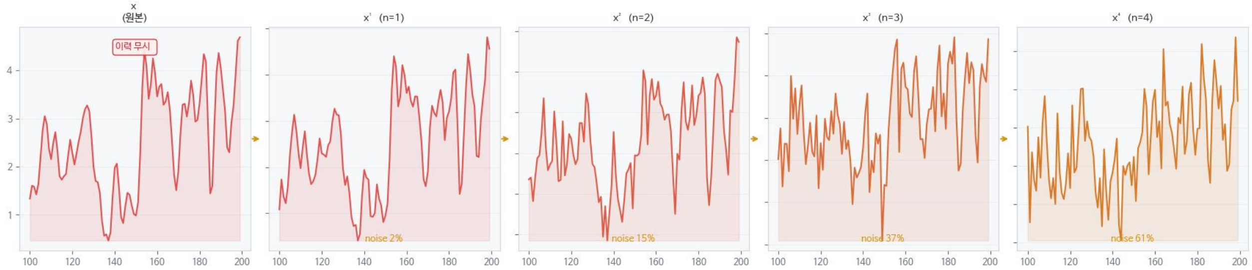
## 연구 배경1. Distribution Drift

- 시계열 데이터는 데이터간 분포의 불일치가 발생
- Diffusion 기반의 기존 방법들은 예측대상이 되는 타겟에만 점진적으로 노이즈를 주입하여 이러한 시계열 패턴을 포착하지 못함

원본 시계열



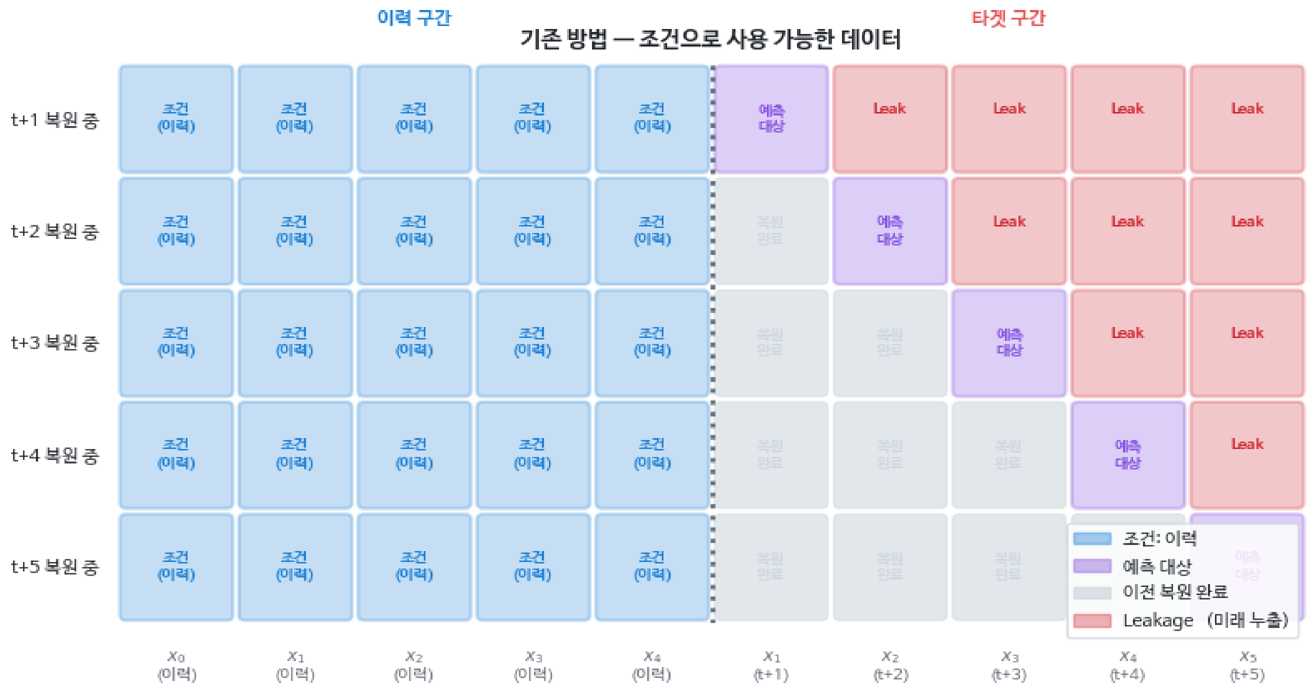

노이즈 주입 과정



# 02 연구 배경

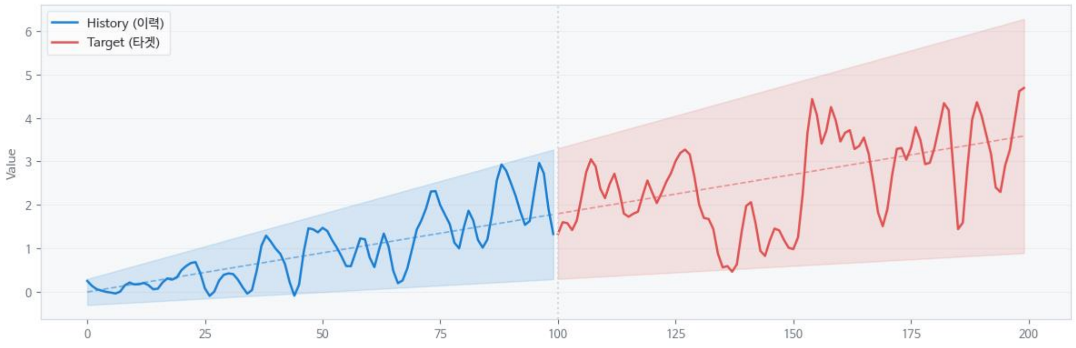
## 연구 배경2. Temporal Causality

- 시계열 데이터는 과거가 미래에 영향을 미침
- Backward process에서는 전체 조건을 한꺼번에 활용하여 노이즈를 제거함



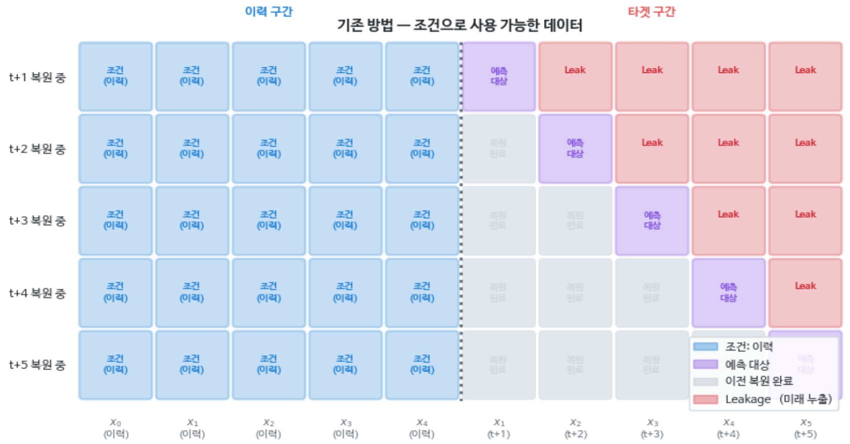
# 02 아이디어

## Distribution Drift




과거 정보를 forward process에 통합하자

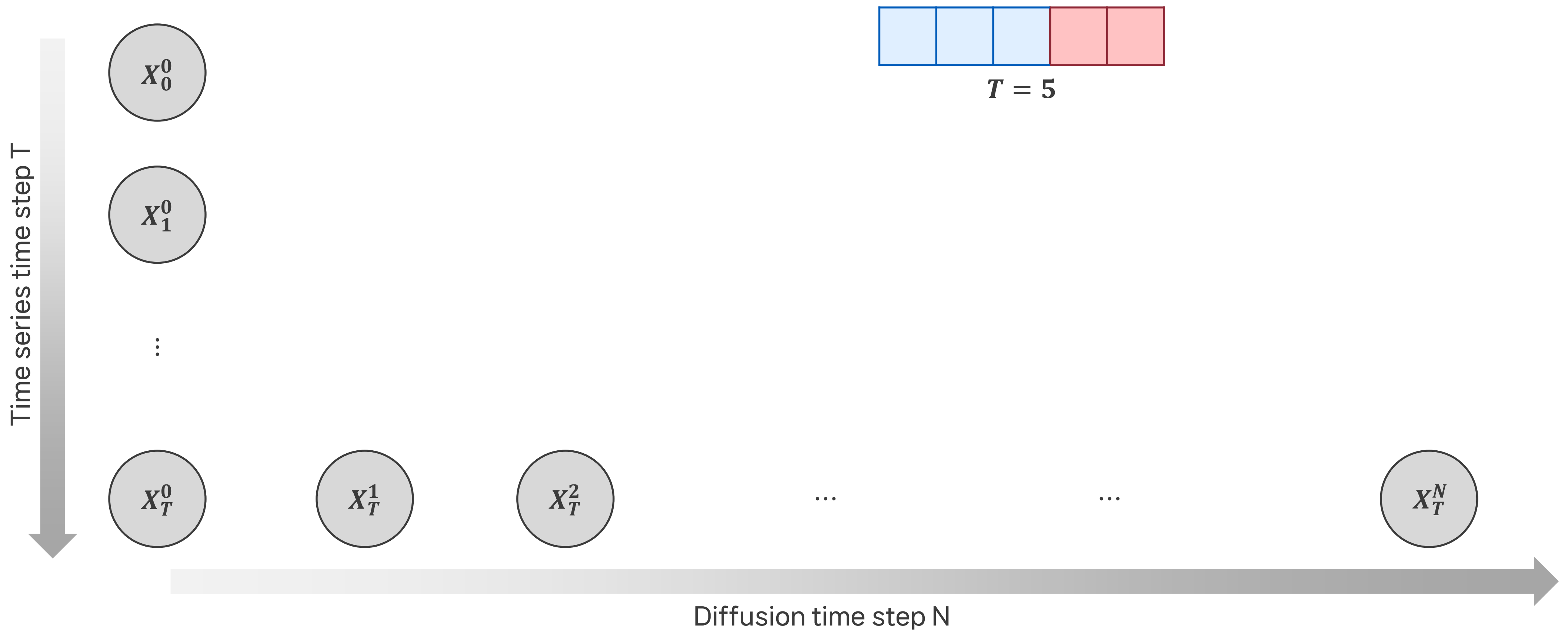
## Temporal Causality




이전 시점의 데이터를 조건으로 하여 복원하자

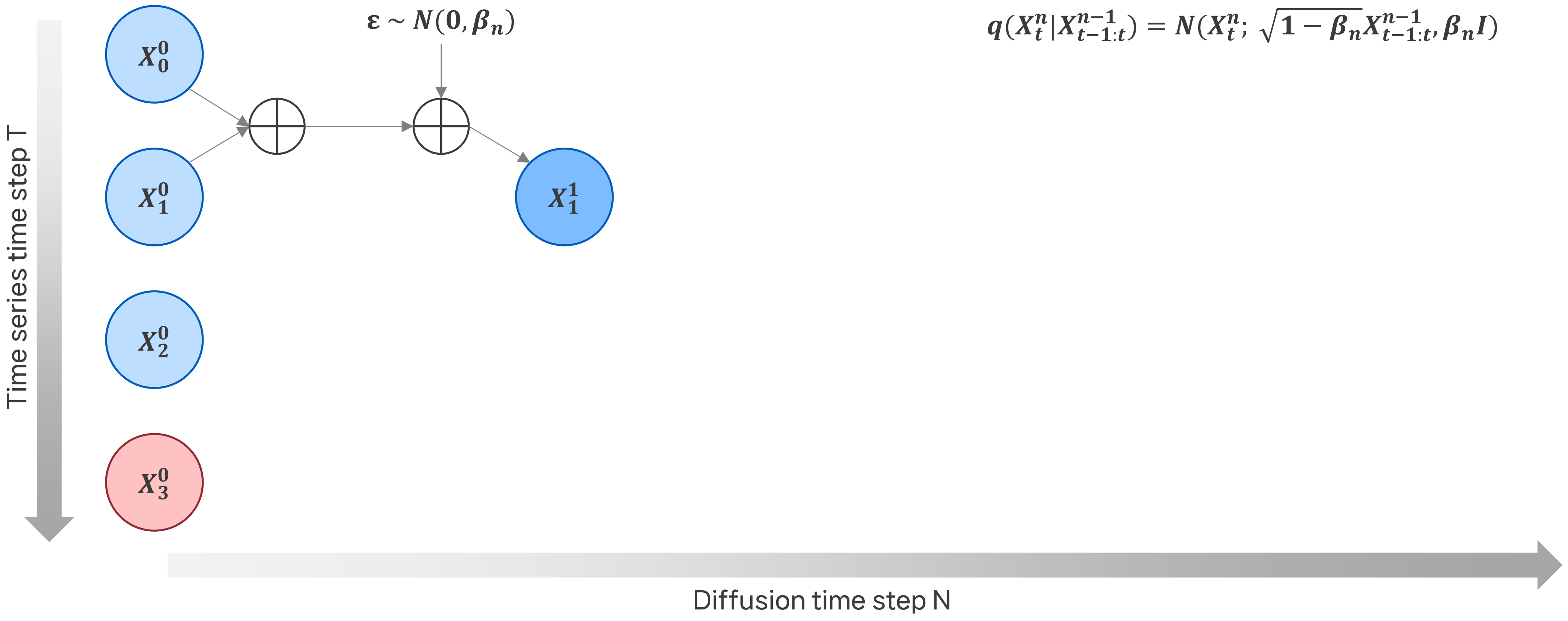
# 02 아이디어

## 아이디어1. Recurrent Forward Diffusion Process



# 02 아이디어

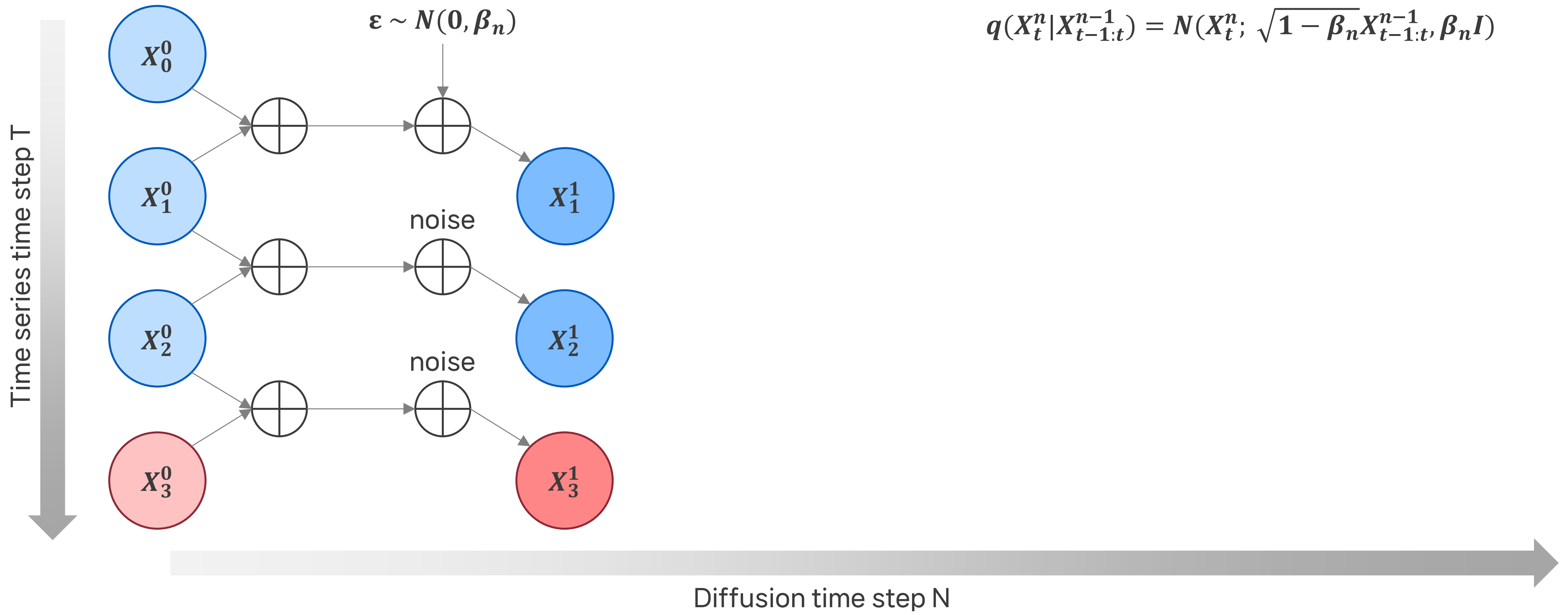
## 아이디어1. Recurrent Forward Diffusion Process




$$q(X_t^n | X_{t-1:t}^{n-1}) = N(X_t^n; \sqrt{1-\beta_n} X_{t-1:t}^{n-1}, \beta_n I)$$

# 02 아이디어

## 아이디어1. Recurrent Forward Diffusion Process

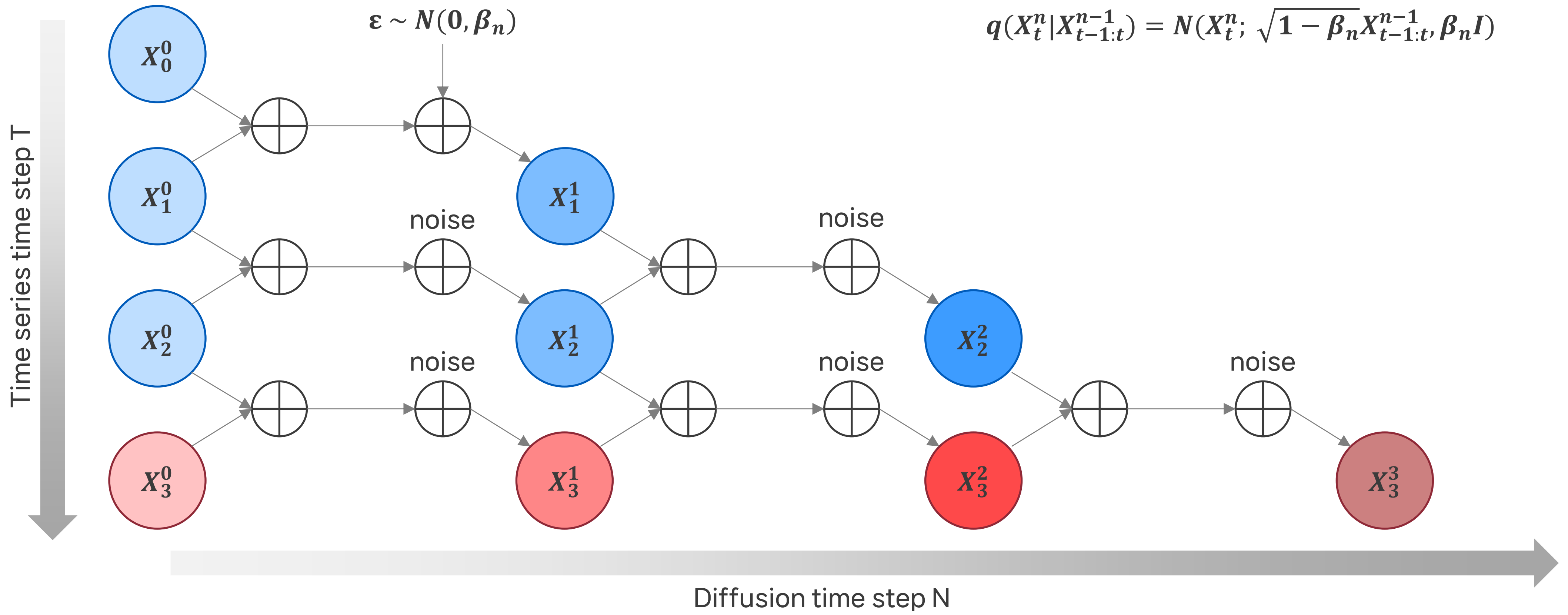



$$q(X_t^n | X_{t-1:t}^{n-1}) = N(X_t^n; \sqrt{1-\beta_n} X_{t-1:t}^{n-1}, \beta_n I)$$

# 02 아이디어

## 아이디어1. Recurrent Forward Diffusion Process

$$q(X_t^n | X_{t-1:t}^{n-1}) = N(X_t^n; \sqrt{1-\beta_n} X_{t-1:t}^{n-1}, \beta_n I)$$

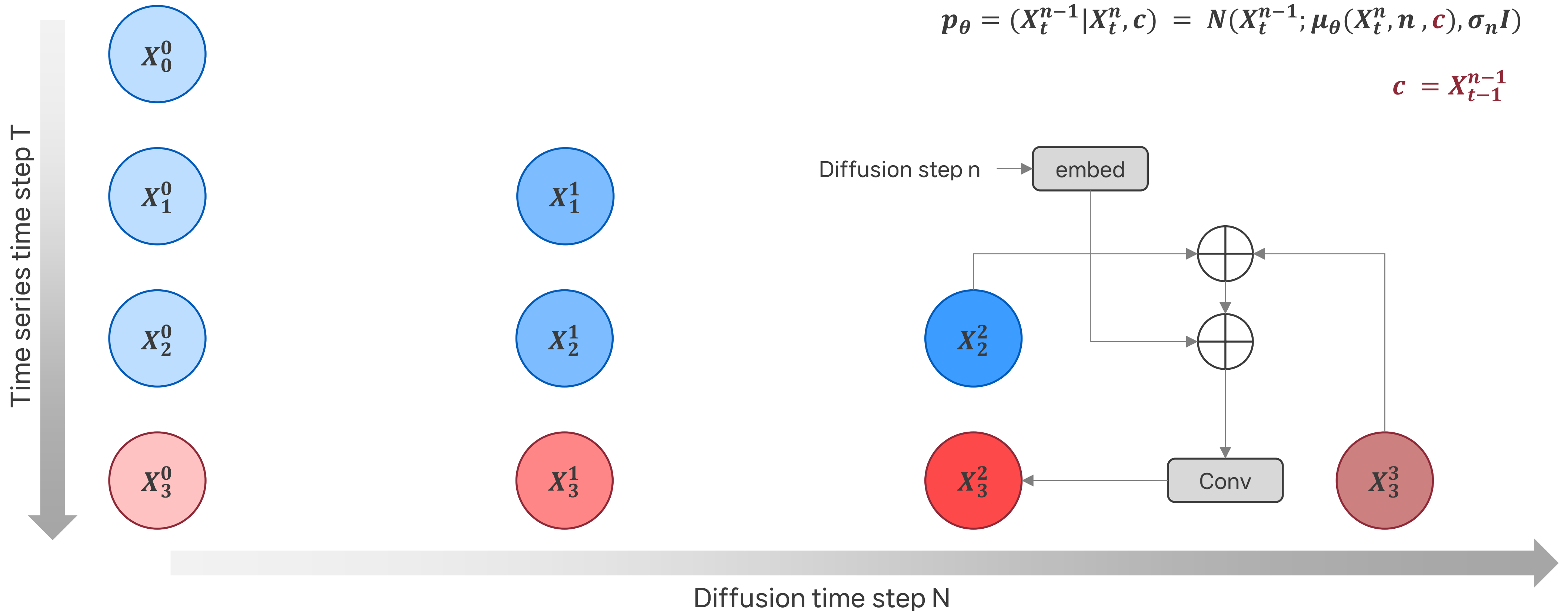


# 02 아이디어

## 아이디어2. Step-Aware Guidance in Backward Process

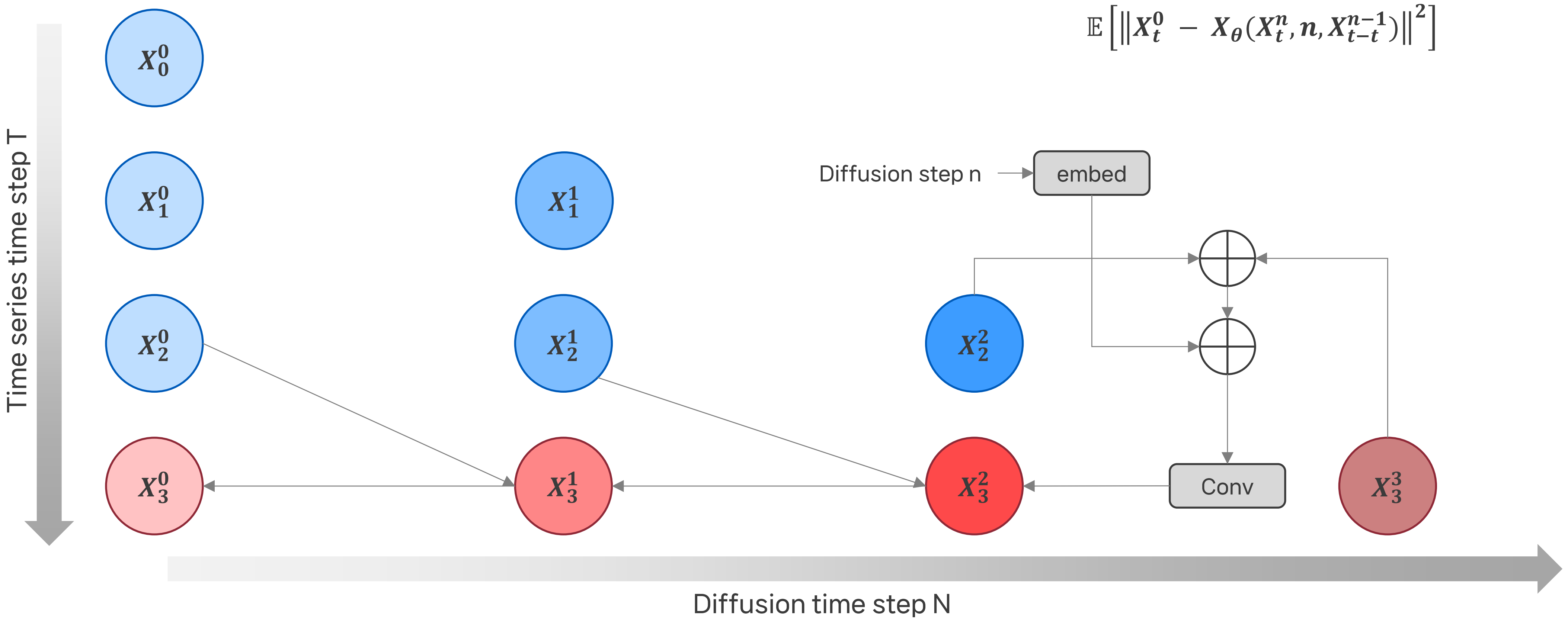
$$p_\theta = (X_t^{n-1} | X_t^n, c) \;=\; N(X_t^{n-1}; \mu_\theta(X_t^n, n, c), \sigma_n I)$$

$$c \;= X_{t-1}^{n-1}$$



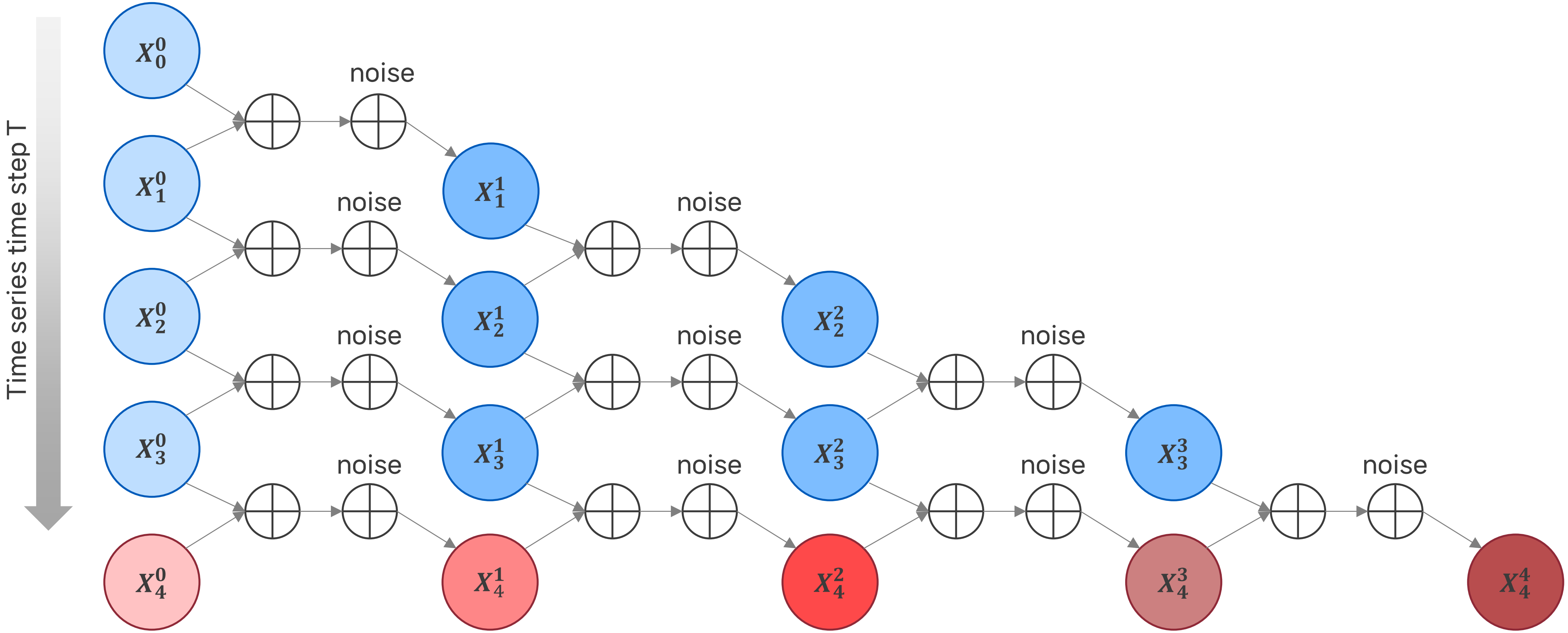
# 02 아이디어

## 아이디어2. Step-Aware Guidance in Backward Process



# 02 아이디어

## 아이디어2. Step-Aware Guidance in Backward Process



# 02 결과

| Model | *Weather* | | *Exchange* | | *M4-Yearly* | | *M4-Quarterly* | | *M4-Monthly* | | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | rank |
| *REDI* | **0.0042** (1) | **0.0072** (1) | 0.0427 (4) | 0.0558 (4) | 0.8104 (2) | **1.2453** (1) | **0.5069** (1) | 0.7077 (2) | **0.5252** (1) | **0.7176** (1) | **1.8** |
| TimeDiff | 0.0070 (4) | 0.0473 (10) | 0.0506 (5) | 0.0654 (5) | 0.8408 (6) | 1.5079 (9) | 0.5810 (2) | **0.6957** (1) | 0.6070 (7) | 0.8014 (2) | 5.1 |
| CSDI | 0.0074 (6) | 0.0215 (9) | 0.1736 (11) | 0.1920 (10) | 0.8650 (7) | 1.5033 (8) | 0.6508 (8) | 0.9233 (6) | 0.6219 (8) | 0.8721 (5) | 7.5 |
| TimeGrad | 0.0075 (7) | 0.0100 (5) | 0.1536 (10) | 0.1883 (9) | 0.8717 (8) | 1.5430 (10) | 0.6539 (9) | 0.9105 (5) | 0.6357 (9) | 0.8617 (3) | 7.5 |
| TransMAF | 0.0070 (4) | 0.0080 (2) | 0.2740(12) | 0.4190 (12) | 0.8762 (9) | 1.4258 (6) | 0.6280 (5) | 0.7321 (3) | 0.7512 (11) | 0.8629 (4) | 6.5 |
| DeepAR | 0.0066 (2) | 0.0086 (3) | 0.1218 (9) | 0.1499 (8) | 0.8827 (10) | 1.4714 (7) | 0.6681 (11) | 0.8990 (4) | 0.7695 (12) | 0.9053 (6) | 7.2 |
| PatchTST | 0.0083 (9) | 0.0120 (7) | **0.0341** (1) | **0.0492** (1) | 0.8368 (4) | 1.3978 (4) | 0.6418 (6) | 1.1500 (9) | 0.5458 (3) | 0.9868 (8) | 4.9 |
| FEDformer | 0.0174 (11) | 0.2681 (12) | 0.0409 (3) | 0.0551 (2) | 0.9807 (12) | 1.6885 (12) | 0.6106 (3) | 1.1140 (7) | 0.5572 (4) | 0.9955 (9) | 7.5 |
| AutoFormer | 0.0483 (12) | 0.0599 (11) | 0.0393 (2) | 0.0549 (3) | 0.8191 (3) | 1.3817 (3) | 0.6557 (10) | 1.1641 (11) | 0.5693 (5) | 1.0012 (10) | 7.0 |
| Informer | 0.0114 (10) | 0.0155 (8) | 0.0816 (7) | 0.4099 (11) | **0.7647** (1) | 1.2860 (2) | 0.6768 (12) | 1.1780 (12) | 0.5877 (6) | 1.0385 (11) | 8.0 |
| Transformer | 0.0069 (3) | 0.0098 (4) | 0.1005 (8) | 0.1402 (7) | 0.8388 (5) | 1.4188 (5) | 0.6278 (4) | 1.1320 (8) | 0.5327 (2) | 0.9729 (7) | 8.0 |
| FiLM | 0.0075 (7) | 0.0118 (6) | 0.0766 (6) | 0.0982 (6) | 0.9551 (11) | 1.6169 (11) | 0.6453 (7) | 1.1577 (10) | 0.6842 (10) | 1.0992 (12) | 8.6 |

# 02 결론

## 기여점

- Diffusion 모델은 사전 분포 가정 없이 데이터 분포를 학습하는 뛰어는 능력으로 시계열 예측에 효과적임
- 하지만 기존 모델들은 distribution drift, temporal causality 문제가 발생하였는데, 본 연구는 recurrent 방식의 forward process 설계와 step-aware guidance를 통해 시계열 예측 성능을 향상시킴

## 고찰

- 윈도우가 커졌을 경우 성능이 좋을 것인가? → 오차 누적의 위험
- Diffusion model의 노이즈 주입 방식을 통해 자연스럽게 최근 데이터에 가중치를 줌

# 03 ARMD

## Auto-Regressive Moving Diffusion Models for Time Series Forecasting (AAAI, 2025)

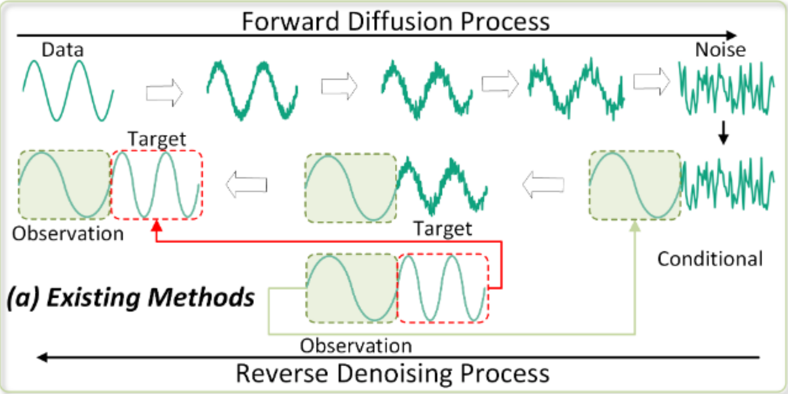
# Auto-Regressive Moving Diffusion Models for Time Series Forecasting

**Jiaxin Gao[1,2*], Qinglong Cao[1,2*], Yuntian Chen[2†]**

[1]Shanghai Jiao Tong University, Shanghai, China
[2]Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, Zhejiang, China
jiaxingao@sjtu.edu.cn; caoql2022@sjtu.edu.cn; ychen@eitech.edu.cn

**Abstract**

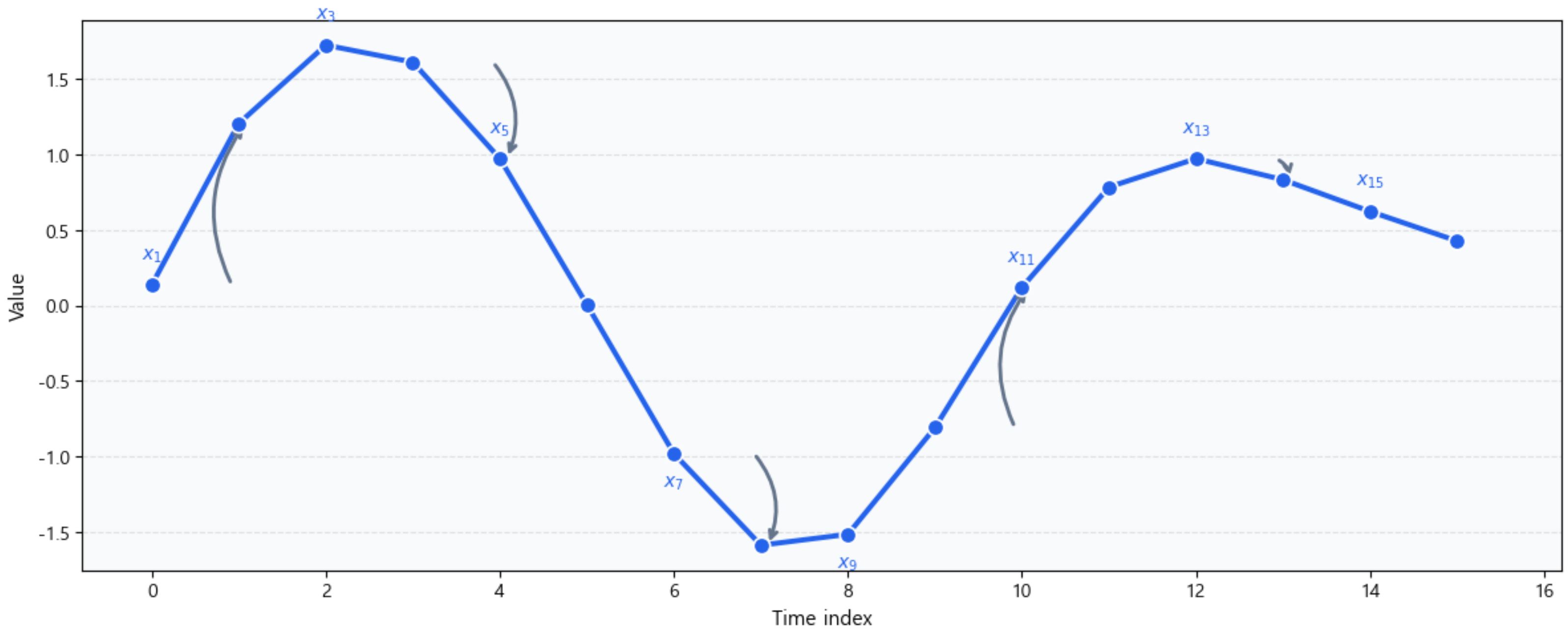
Time series forecasting (TSF) is essential in various domains, and recent advancements in diffusion-based TSF models have shown considerable promise. However, these models typically adopt traditional diffusion patterns, treating TSF as a noise-based conditional generation task. This approach neglects the inherent continuous sequential nature of time series, leading to a fundamental misalignment between diffusion mechanisms and the TSF objective, thereby severely impairing performance. To bridge this misalignment, and inspired by the classic Auto-Regressive Moving Average



# 03 연구 배경

## 연구 배경

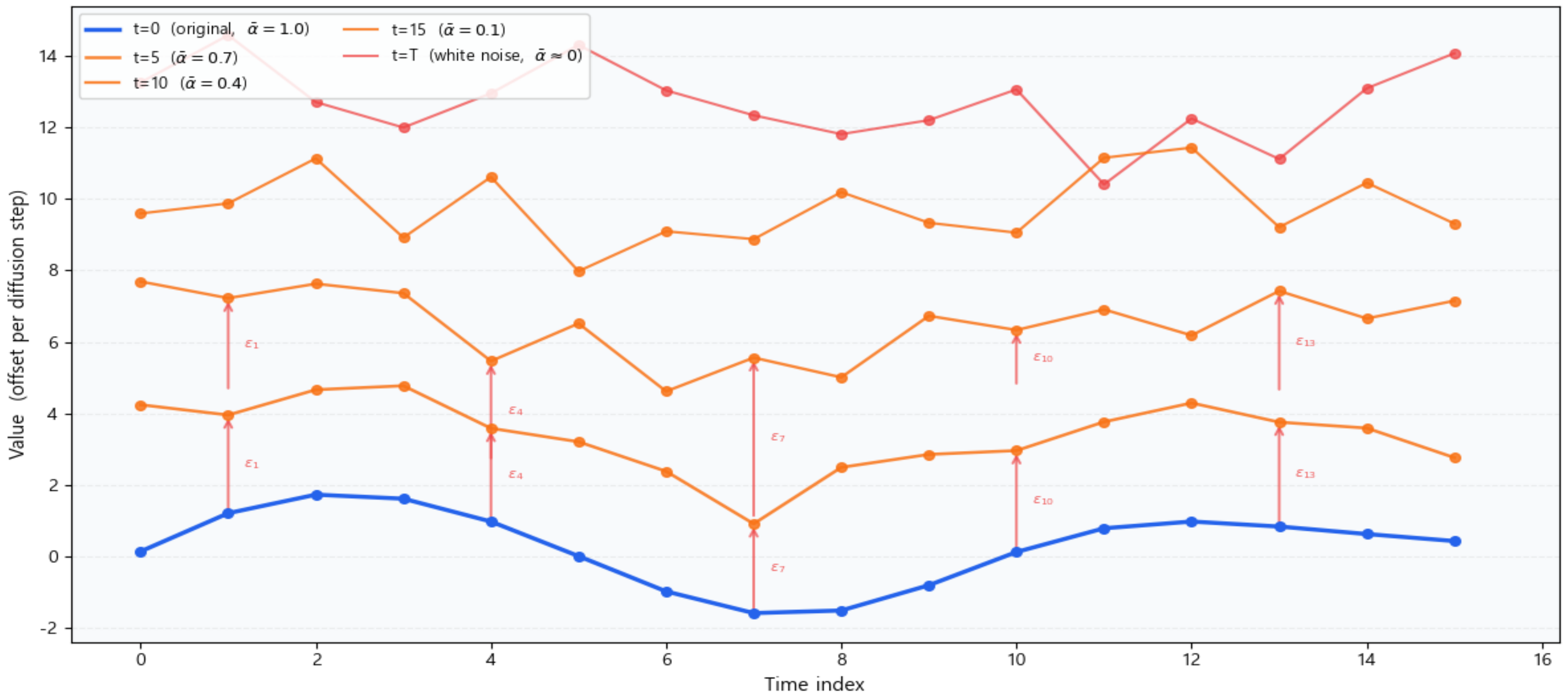
- 시계열 데이터에 노이즈를 주입하는 과정은 시계열의 연속적 순차 특성을 간과함
- 시계열 데이터에 노이즈가 주입되면서, 시계열 진화 과정에 담긴 귀중한 중간 정보를 포착하지 못하고 활용하지 못함



# 03 연구 배경

## 연구 배경

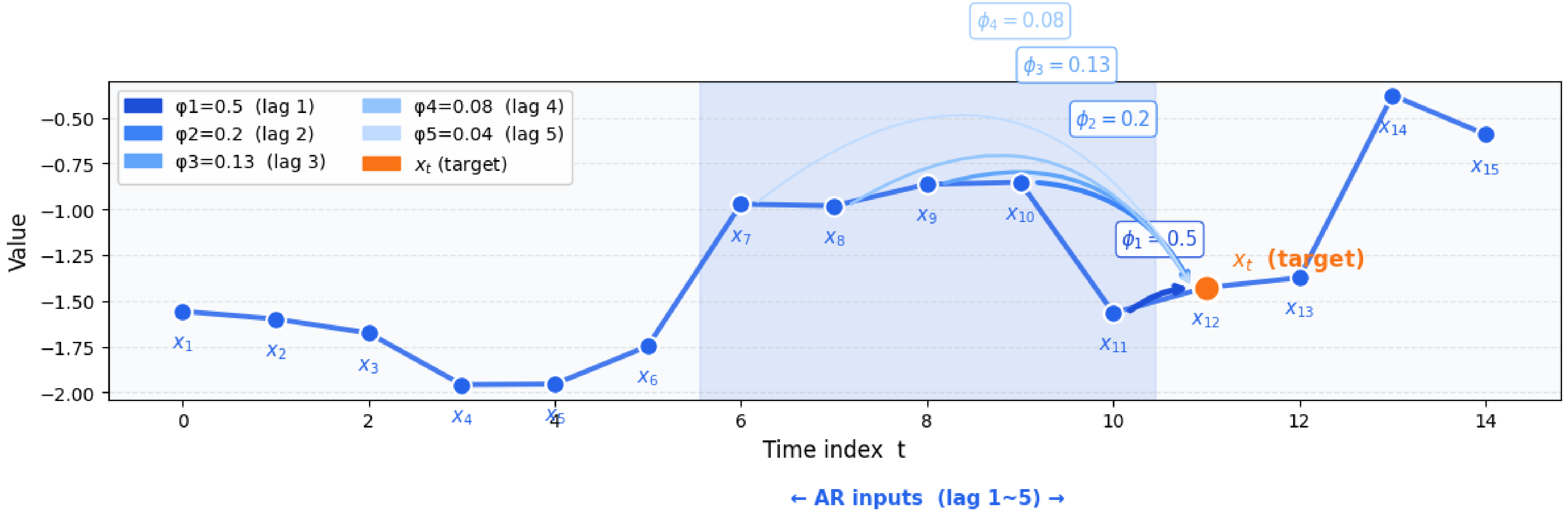
- 시계열 데이터에 노이즈를 주입하는 과정은 시계열의 연속적 순차 특성을 간과함
- 시계열 데이터에 노이즈가 주입되면서, 시계열 진화 과정에 담긴 귀중한 중간 정보를 포착하지 못하고 활용하지 못함



## ARMA (Auto-Regressive Moving Average)

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots \phi_p x_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

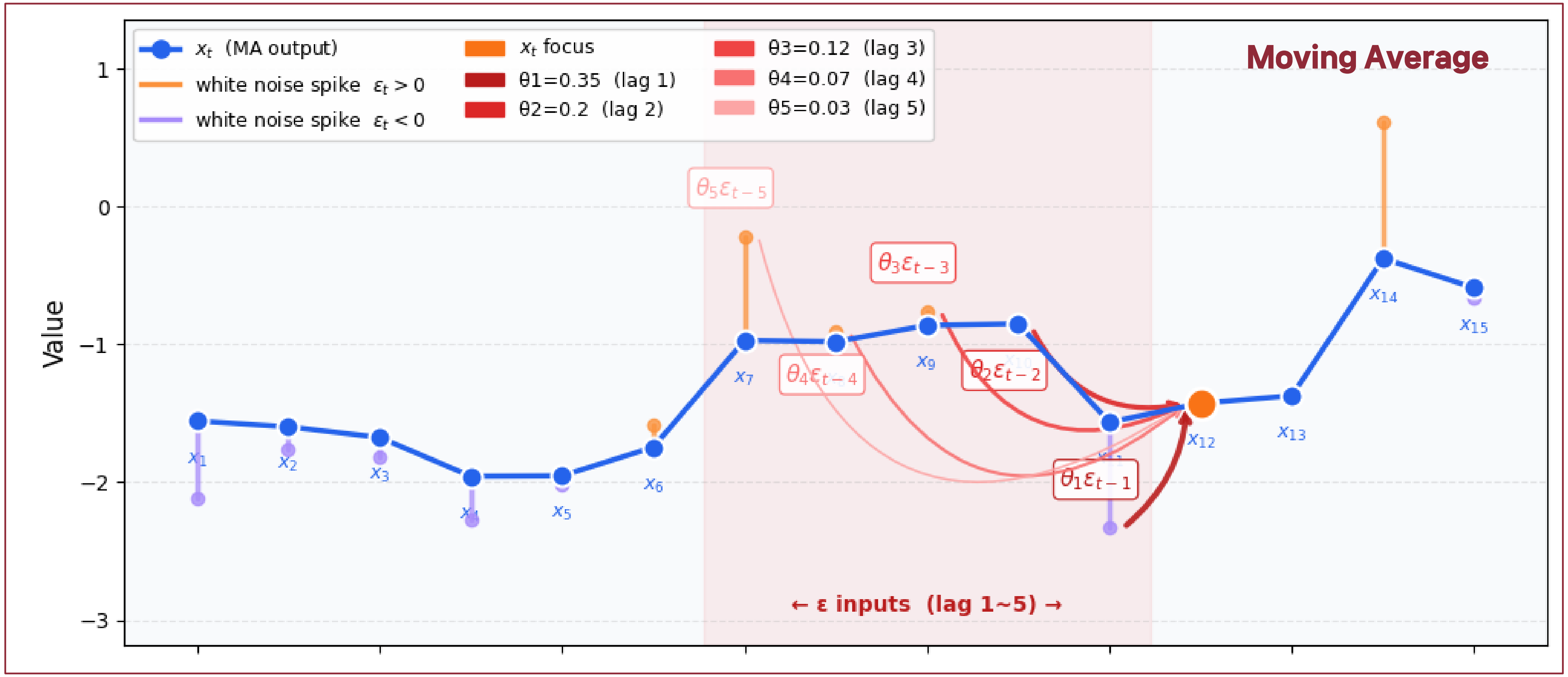
### Auto-Regressive



# 03 아이디어

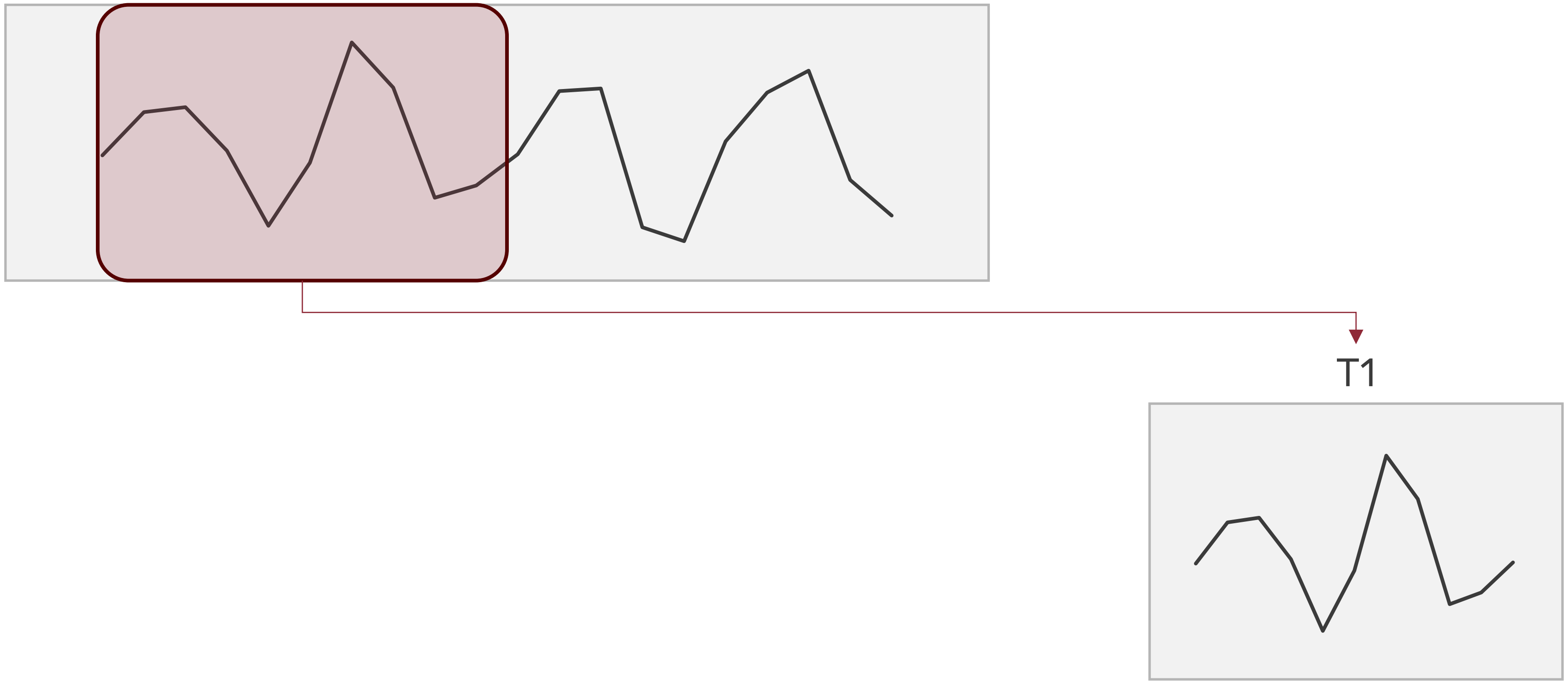
## ARMA (Auto-Regressive Moving Average)

$$x_t \;=\; \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots \phi_p x_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$



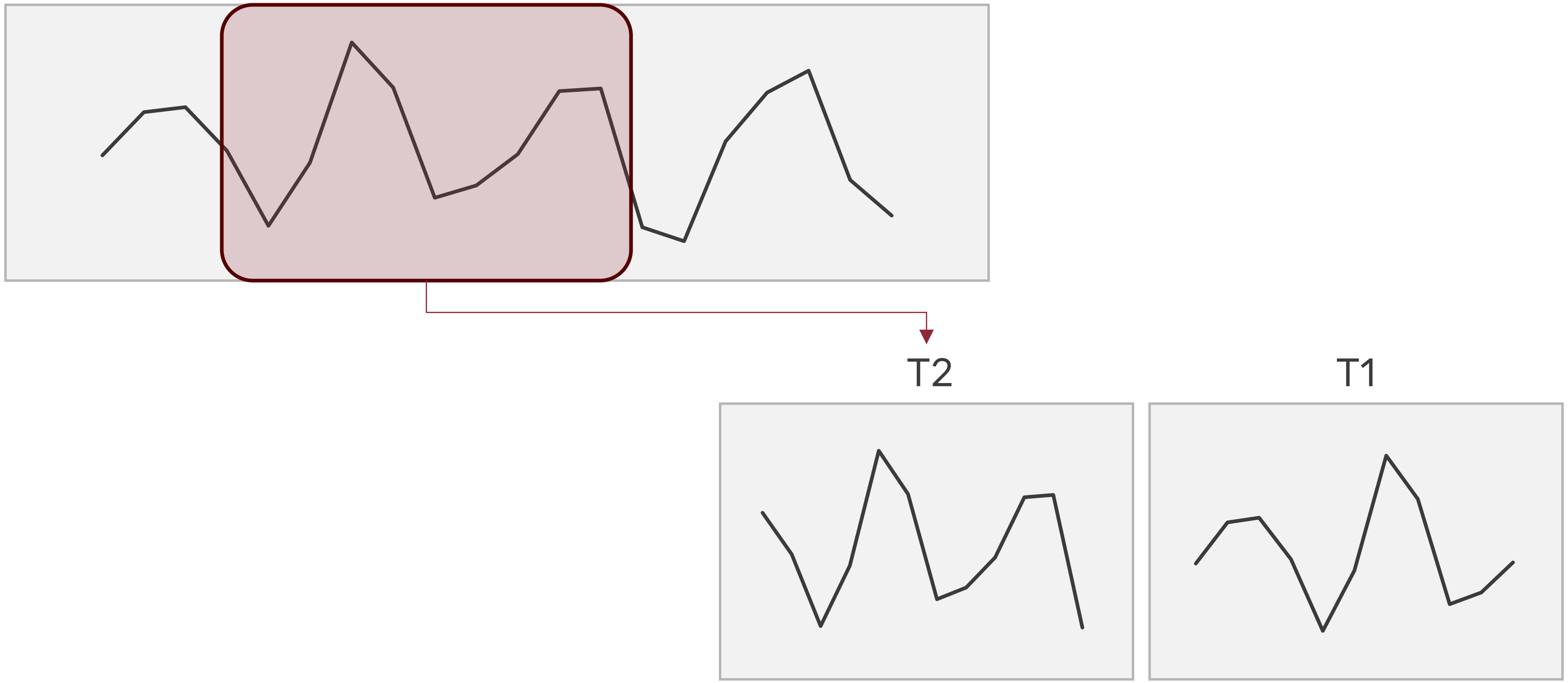
# 03 아이디어

## Forward process의 재설계



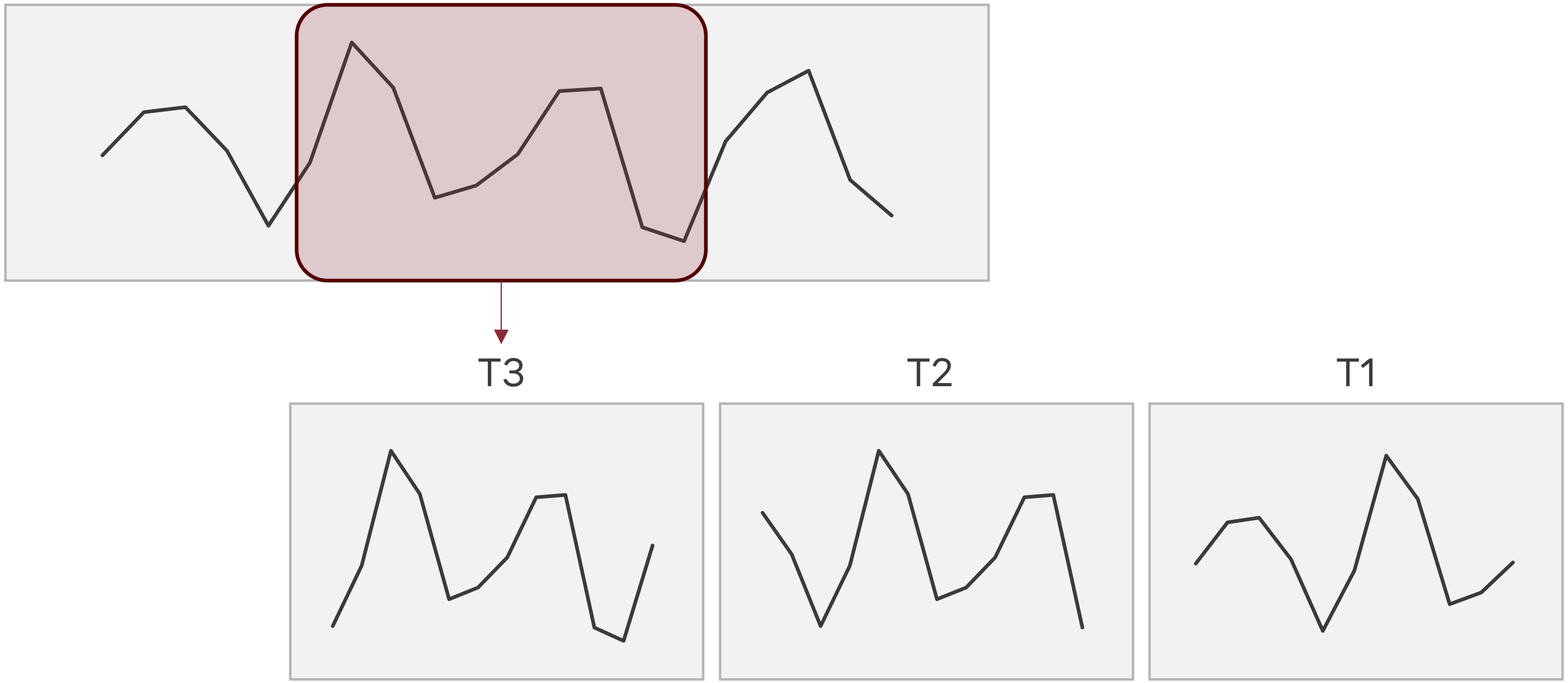
# 03 아이디어

## Forward process의 재설계



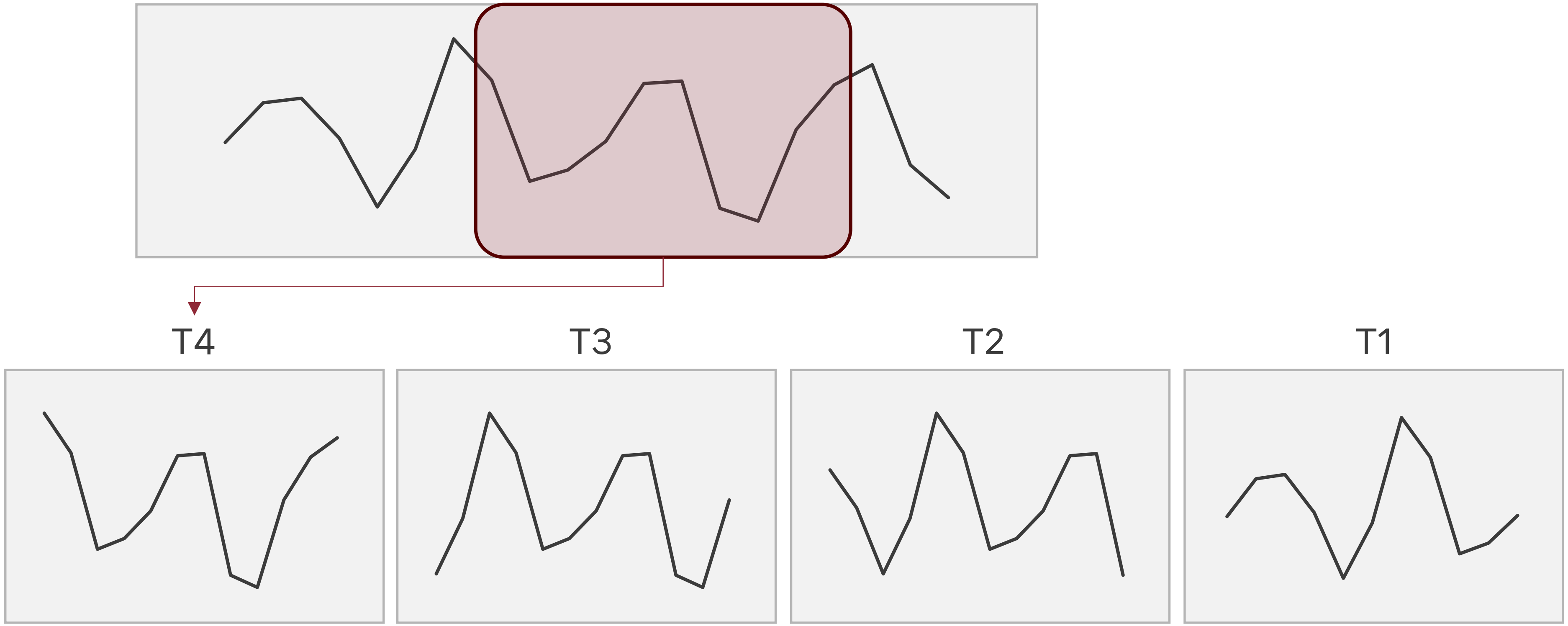
# 03 아이디어

## Forward process의 재설계



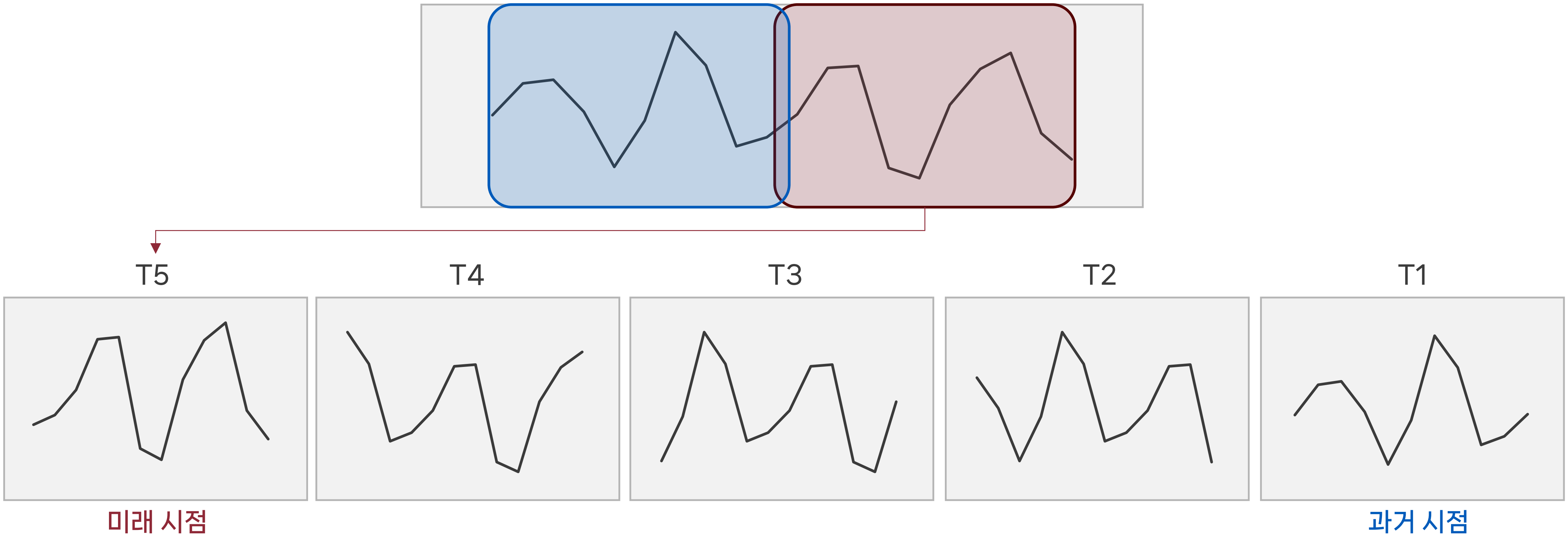
# 03 아이디어

## Forward process의 재설계



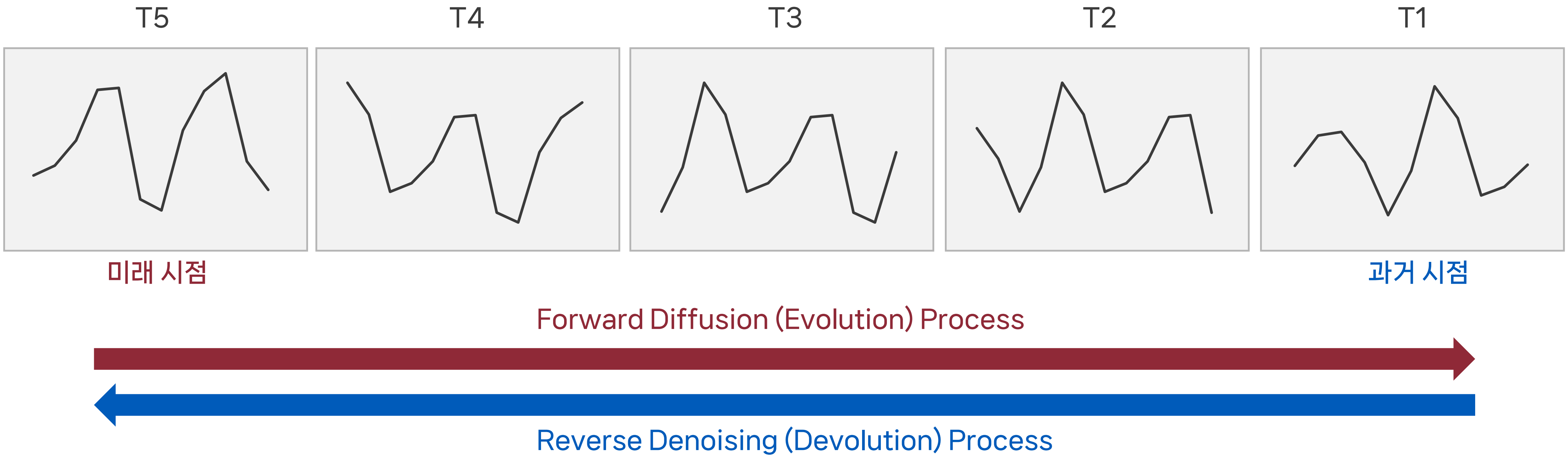
# 03 아이디어

## Forward process의 재설계



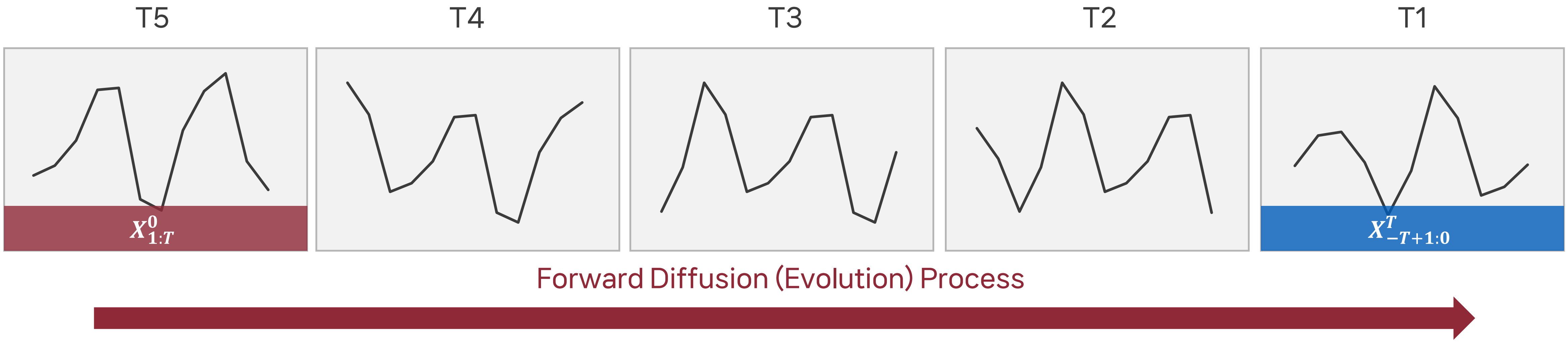
# 03 아이디어

## Forward process의 재설계



# 03 아이디어

## Forward process의 재설계



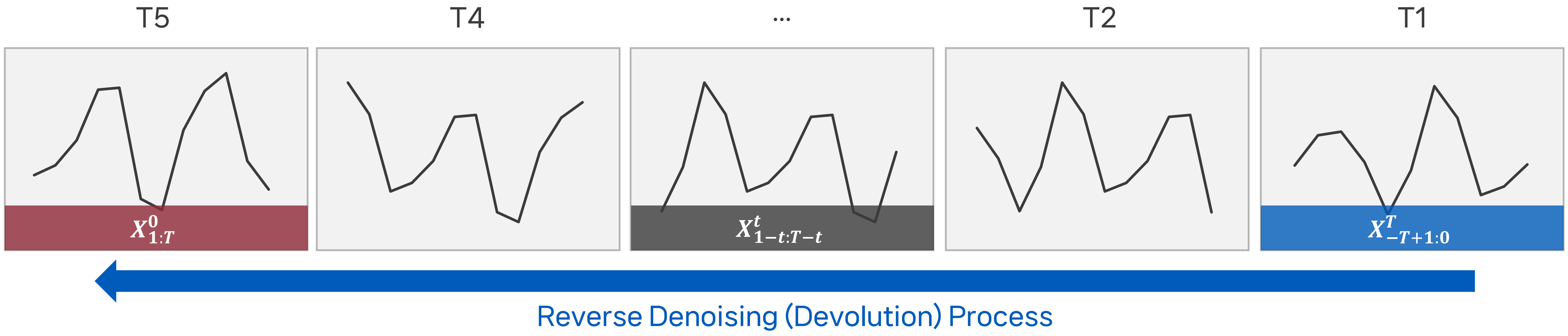

$$X^t_{1-t:T-t} = Slide(X^{t-1}_{2-t:T-t+1}, 1)$$

$$X^t_{1-t:T-t} = Slide(X^0_{1:T}, t) = \sqrt{\bar{\alpha}_t} X^0_{1:T} + \sqrt{1-\bar{\alpha}_t} z_t$$

$$z^t = \left( \sqrt{\frac{1}{\bar{\alpha}_t}} X^t_{1-t:T-t} - X^0_{1:T} \right) / \sqrt{\frac{1}{\bar{\alpha}_t} - 1}$$

# 03 아이디어

## Reverse process 재설계




$$\hat{X}^0(X^t,t,\theta) = \frac{W(t)\cdot X^t_{1-t:T-t} + (1-bW(t))\cdot D}{(1+cW(t))^d}$$

$$\hat{z}(t,\theta) = \left(\sqrt{\frac{1}{\bar{\alpha}_t}}X^t_{1-t:T-t} - \hat{X}^0(X^t,t,\theta)\right) / \sqrt{\frac{1}{\bar{\alpha}_t} - 1}$$

# 03 결과

| Methods | Metric | Solar Energy | ETTh1 | ETTh2 | ETTm1 | ETTm2 | Exchange | Stock | Best Count |
|---|---|---|---|---|---|---|---|---|---|
| **ARMD (Ours)** | MSE | **0.167** | **0.445** | 0.311 | **0.337** | **0.181** | **0.093** | **0.235** | **12** |
| | MAE | **0.236** | **0.459** | **0.338** | **0.376** | 0.255 | **0.203** | **0.269** | |
| Diffusion-TS (Yuan and Qiao 2024) | MSE | 0.181 | 0.643 | 0.544 | 0.678 | 0.497 | 0.275 | 0.416 | 0 |
| | MAE | 0.252 | 0.586 | 0.494 | 0.613 | 0.459 | 0.382 | 0.533 | |
| MG-TSD (Fan et al. 2024) | MSE | 0.443 | 1.096 | **0.295** | 0.690 | 0.202 | 0.396 | 0.365 | 1 |
| | MAE | 0.529 | 0.765 | 0.345 | 0.631 | 0.278 | 0.460 | 0.453 | |
| TSDiff (Kollovieh et al. 2024) | MSE | 0.352 | 0.614 | 0.470 | 0.686 | 0.242 | 0.125 | 0.330 | 0 |
| | MAE | 0.432 | 0.521 | 0.418 | 0.603 | 0.311 | 0.240 | 0.365 | |
| D3VAE (Li et al. 2022) | MSE | 0.416 | 1.123 | 0.389 | 0.644 | 0.394 | 0.240 | 0.345 | 0 |
| | MAE | 0.492 | 0.728 | 0.373 | 0.538 | 0.410 | 0.371 | 0.390 | |
| TimeGrad (Rasul et al. 2021) | MSE | 0.359 | 0.884 | 0.297 | 0.661 | 0.182 | 0.508 | 0.333 | 1 |
| | MAE | 0.449 | 0.725 | 0.349 | 0.639 | **0.254** | 0.554 | 0.376 | |

# 03 결과

| Methods | Metric | Solar Energy | ETTh1 | ETTh2 | ETTm1 | ETTm2 | Exchange | Stock | Best Count |
|---|---|---|---|---|---|---|---|---|---|
| **ARMD (Ours)** | MSE | **0.167** | 0.445 | 0.311 | 0.337 | 0.181 | 0.093 | **0.235** | **7** |
| | MAE | **0.236** | 0.459 | **0.338** | 0.376 | **0.255** | **0.203** | **0.269** | |
| iTransformer (Liu et al. 2024) | MSE | 0.203 | 0.386 | **0.297** | 0.334 | 0.180 | **0.086** | 0.342 | 2 |
| | MAE | 0.237 | 0.405 | 0.349 | 0.368 | 0.264 | 0.206 | 0.413 | |
| TimesNet (Wu et al. 2023) | MSE | 0.250 | **0.384** | 0.340 | 0.338 | 0.187 | 0.107 | 0.427 | 1 |
| | MAE | 0.292 | 0.402 | 0.347 | 0.375 | 0.267 | 0.234 | 0.499 | |
| DLinear (Zeng et al. 2023) | MSE | 0.290 | 0.386 | 0.333 | 0.345 | 0.193 | 0.088 | 0.286 | 1 |
| | MAE | 0.378 | **0.400** | 0.387 | 0.372 | 0.292 | 0.218 | 0.325 | |
| PatchTST (Nie et al. 2022) | MSE | 0.234 | 0.414 | 0.302 | **0.329** | **0.175** | 0.088 | 0.516 | 3 |
| | MAE | 0.286 | 0.419 | 0.348 | **0.367** | 0.259 | 0.205 | 0.524 | |
| Client (Gao, Hu, and Chen 2023) | MSE | 0.199 | 0.392 | 0.305 | 0.336 | 0.184 | **0.086** | 0.352 | 1 |
| | MAE | 0.239 | 0.409 | 0.353 | 0.369 | 0.267 | 0.206 | 0.433 | |

# 03 결론

## 기여점

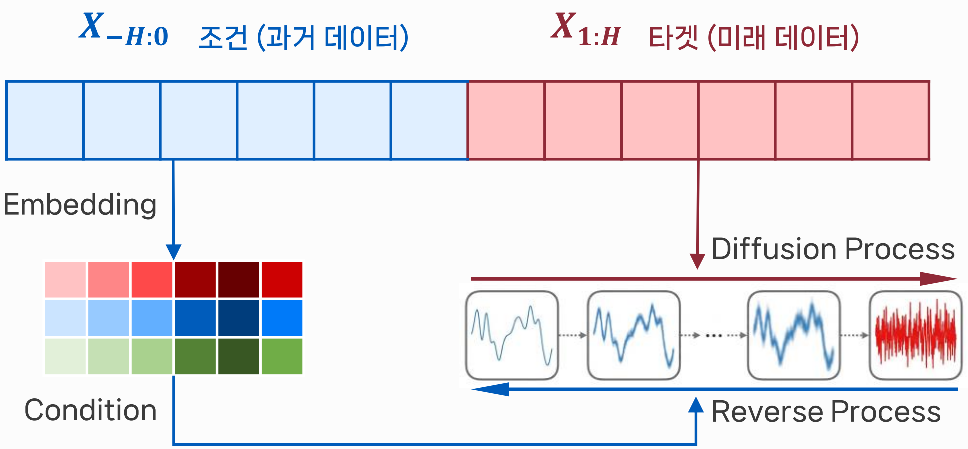
- 최초의 연속 순차 diffusion 기반 시계열 예측 모델로, 기존 diffusion mechanism과 시계열 예측의 목표 불일치를 해소
- 슬라이딩 윈도우로 중간 상태를 생성함으로써, 모든 중간 상태가 시계열 데이터에서 직접 추출된 관찰 가능한 구간

## 고찰

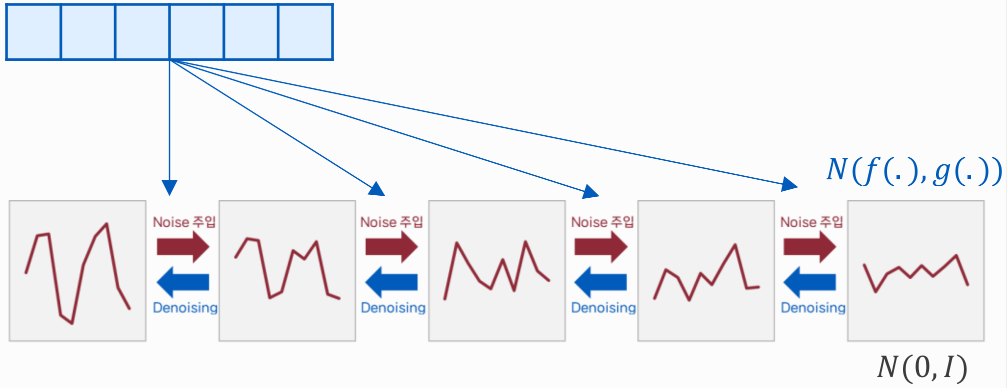
- 예측 길이와 과거 길이가 반드시 같아야 함
- 확률적 예측에 대한 정체성 모호

# Diffusion models for time series forecasting

## Feature-centric




## Diffusion-centric



# 고맙습니다

# 참고 문헌

[1] Gao, J., Cao, Q., & Chen, Y. (2025, April). Auto-regressive moving diffusion models for time series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 39, No. 16, pp. 16727-16735).

[2] Zhou, S., Gu, Z., Xiong, Y., Luo, Y., Wang, Q., & Gao, X. (2024, October). Redi: Recurrent diffusion model for probabilistic time series forecasting. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management* (pp. 3505-3514).