
Feature-centric Diffusion Models For Time Series Forecasting

DMQA Open Seminar

2025.08.29

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

- Background

- Diffusion Models for Time Series Forecasting (TSF)
- TSF Diffusion modeling
- Feature-centric models

02

- Feature 강화 모델

- mr-Diff
- 방법론
- Conclusion

03

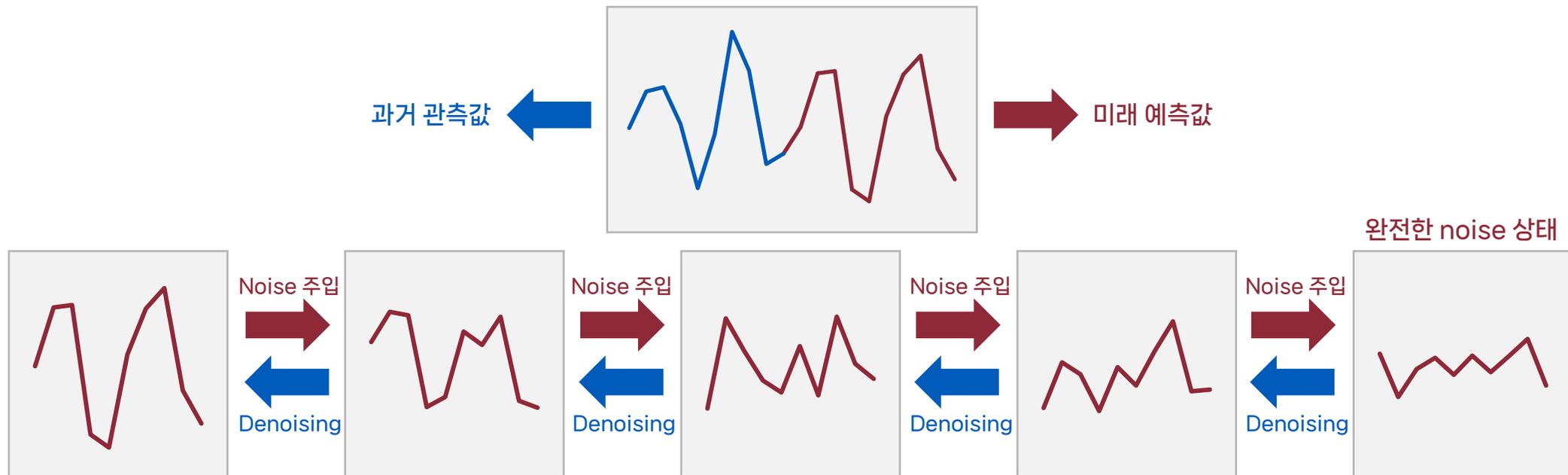
- 불확실성 모델

- D3U
- 방법론
- Conclusion

01 Diffusion Models for Time Series Forecasting (TSF)

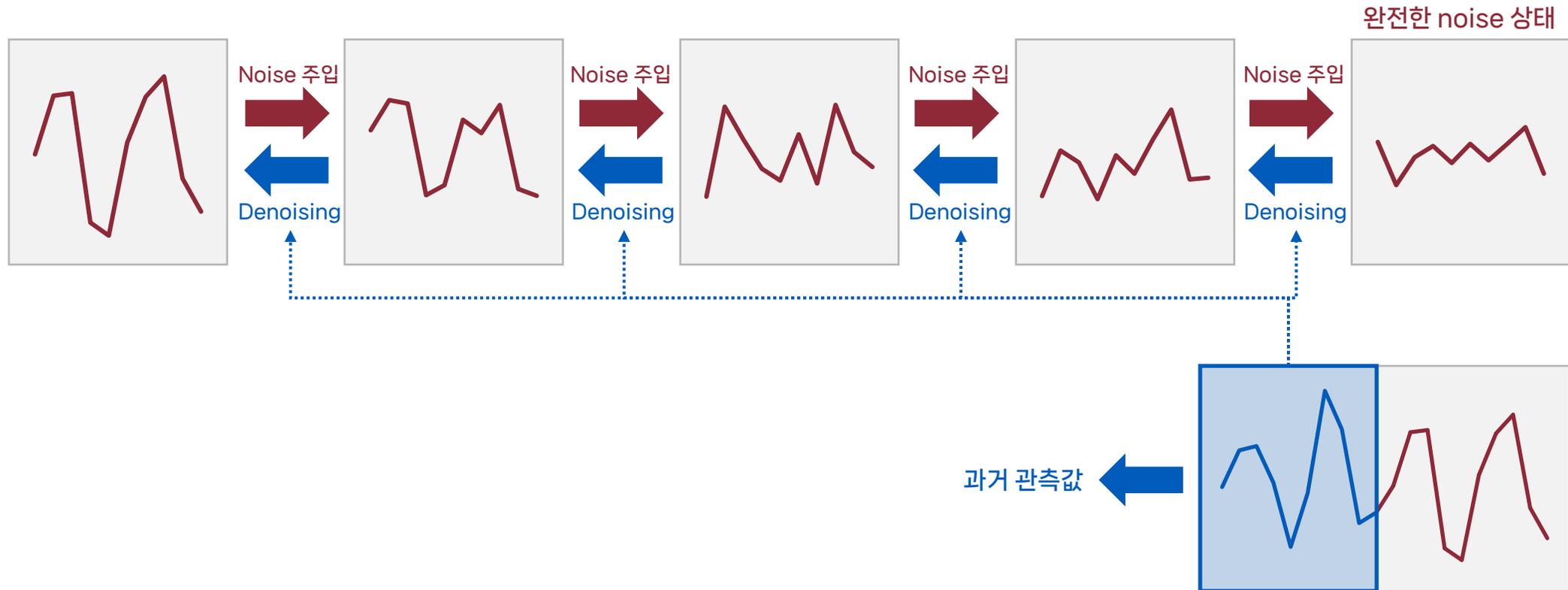
Diffusion Models for TSF

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



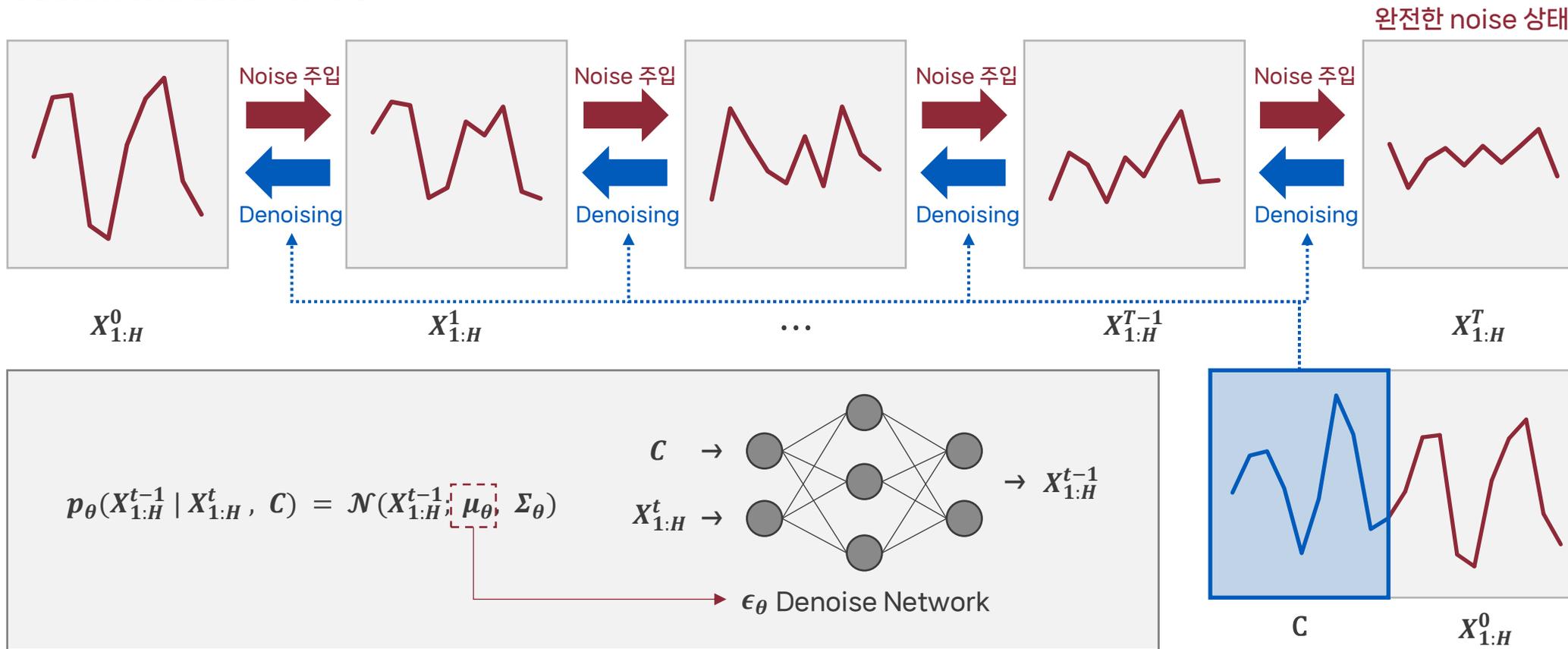
01 Diffusion Models for Time Series Forecasting (TSF)

Diffusion Models for TSF



01 Diffusion Models for Time Series Forecasting (TSF)

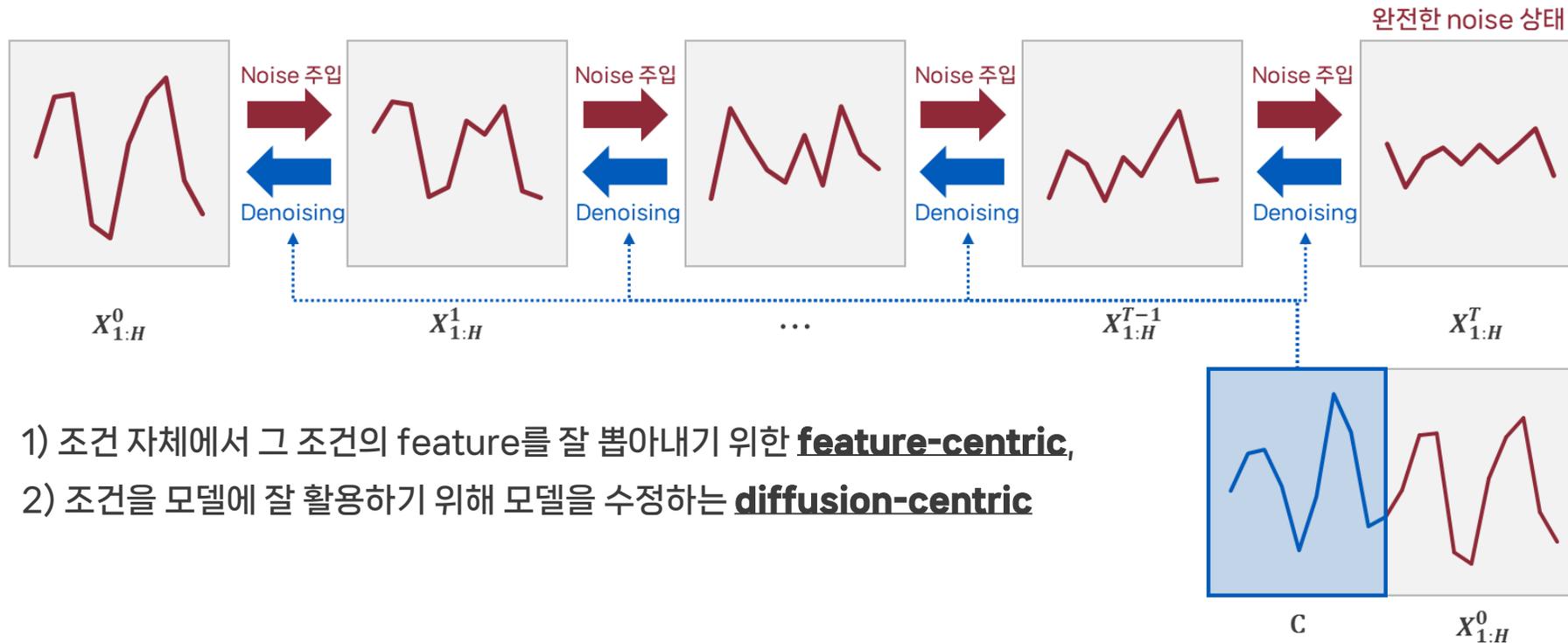
Diffusion Models for TSF



01 TSF Diffusion modeling

Diffusion 기반 시계열 데이터 모델의 핵심

- 본질적으로 "생성" 문제 이므로, 어떤 정보를 조건으로 사용할 것인지, 그리고 어떻게 활용한 것인지가 성능에 큰 영향을 줌

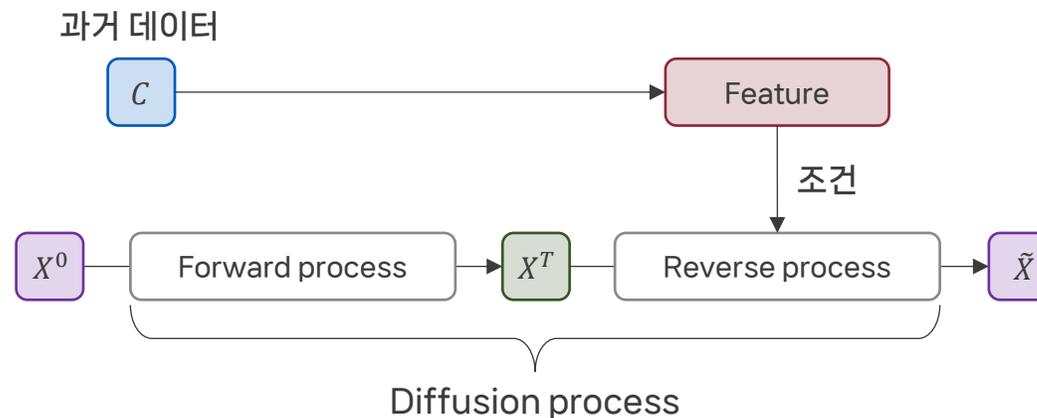
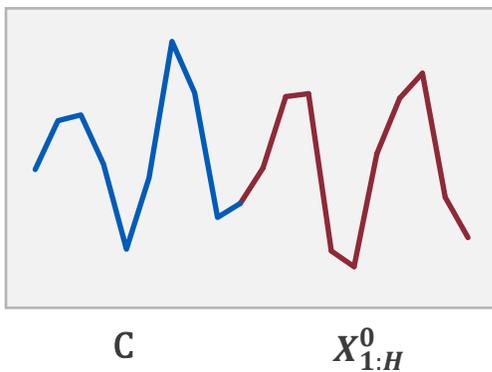


- 조건 자체에서 그 조건의 feature를 잘 뽑아내기 위한 **feature-centric**,
- 조건을 모델에 잘 활용하기 위해 모델을 수정하는 **diffusion-centric**

01 Feature-centric models

Feature-centric model의 종류

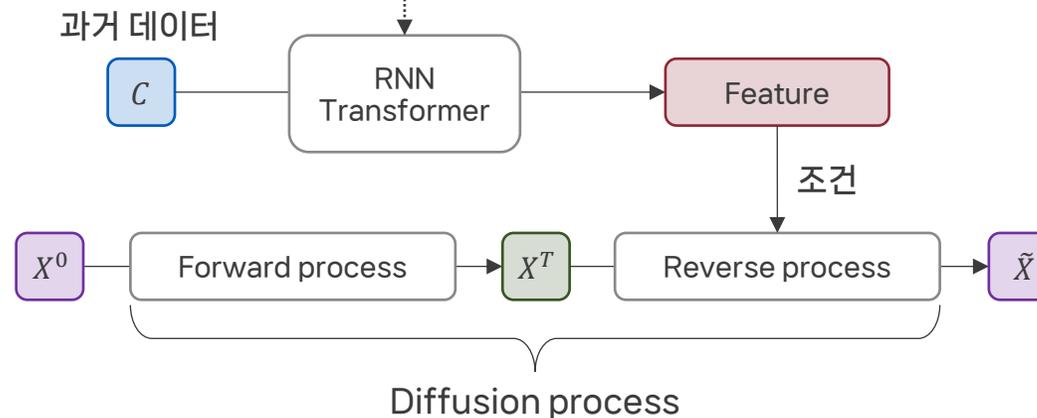
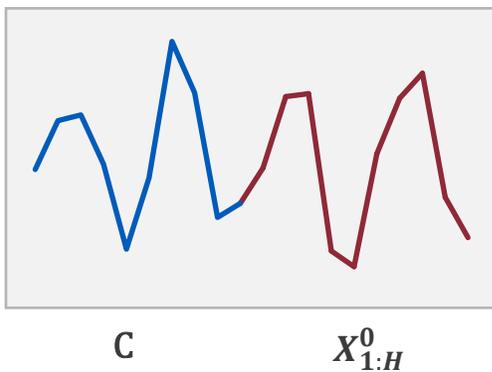
- 조건 자체에서 그 조건의 feature를 잘 뽑아내기 위한 방법을 연구
- RNN 기반 모델, Transformer 기반 모델, Feature 강화 모델, 불확실성 모델



01 Feature-centric models

Feature-centric model의 종류

- 조건 자체에서 그 조건의 feature를 잘 뽑아내기 위한 방법을 연구
- RNN 기반 모델, Transformer 기반 모델, Feature 강화 모델, 불확실성 모델



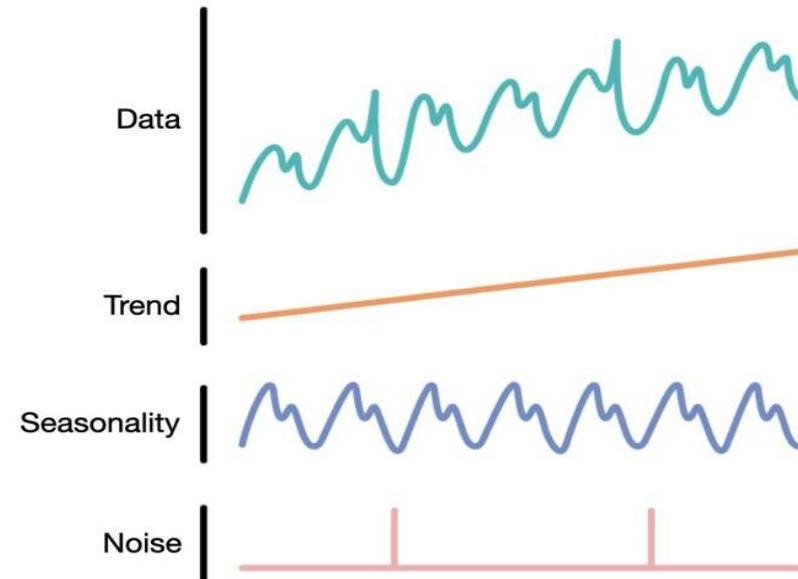
01 Feature-centric models

Feature-centric model의 종류

- 조건 자체에서 그 조건의 feature를 잘 뽑아내기 위한 방법을 연구
- RNN 기반 모델, Transformer 기반 모델, Feature 강화 모델, 불확실성 모델

Component of Time Series

- Trend (추세)
 - 데이터의 장기적인 방향성
 - 시간이 지남에 따라 증가하거나 감소하는 패턴
- Seasonality (계절성)
 - 일정한 주기로 반복되는 패턴 (아이스크림, 전력 소비)
- Noise (불규칙 변동)
 - 예측할 수 없는 무작위적 변동





02 mr-Diff

mr-Diff: Multi-Resolution Diffusion Models for Time Series Forecasting (ICLR, 2024 poster)

- 제목에서 알 수 있듯이, Multi-Resolution을 사용하여 feature 강화

Multi-Resolution

- 시계열은 여러 scale의 패턴을 보이므로, 그, multi-resolution 구조를 활용
- 서로 다른 샘플링 scale의 시계열은 본질적으로 다른 특성을 보임
- 미세한 scale에서는 상세한 패턴을 보여주고, 거시적 scale에서는 전체적인 변동을 보여줌

Multi-Resolution Diffusion Models for Time Series Forecasting



Lifeng Shen, Weiyu Chen, James Kwok

Published: 16 Jan 2024, Last Modified: 24 Mar 2024 ICLR 2024 poster Everyone Revisions BibTeX

Code Of Ethics: I acknowledge that I and all co-authors of this work have read and commit to adhering to the ICLR Code of Ethics.

Keywords: diffusion model, time series, multiscale

Submission Guidelines: I certify that this submission complies with the submission instructions as described on <https://iclr.cc/Conferences/2024/AuthorGuide>.

Abstract:

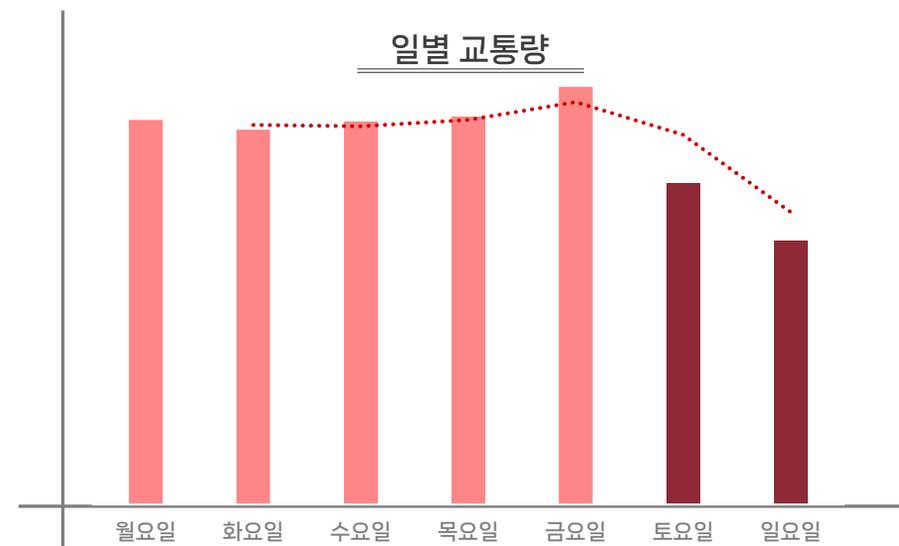
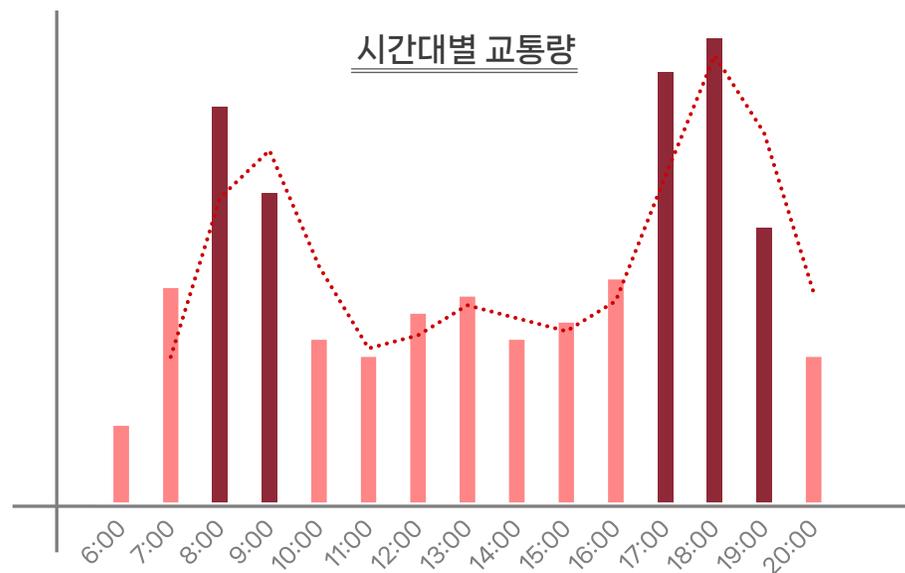
The diffusion model has been successfully used in many computer vision applications, such as text-guided image generation and image-to-image translation. Recently, there have been attempts on extending the diffusion model for time series data. However, these extensions are fairly straightforward and do not utilize the unique properties of time series data. As different patterns are usually exhibited at multiple scales of a time series, we in this paper leverage this multi-resolution temporal structure and propose the multi-resolution diffusion model (mr-Diff). By using the seasonal-trend decomposition, we sequentially extract fine-to-coarse trends from the time series for forward diffusion. The denoising process then proceeds in an easy-to-hard non-autoregressive manner. The coarsest trend is generated first. Finer details are progressively added, using the predicted coarser trends as condition variables. Experimental results on nine real-world time series datasets demonstrate that mr-Diff outperforms state-of-the-art time series diffusion models. It is also better than or comparable across a wide variety of advanced time series prediction models.



02 mr-Diff

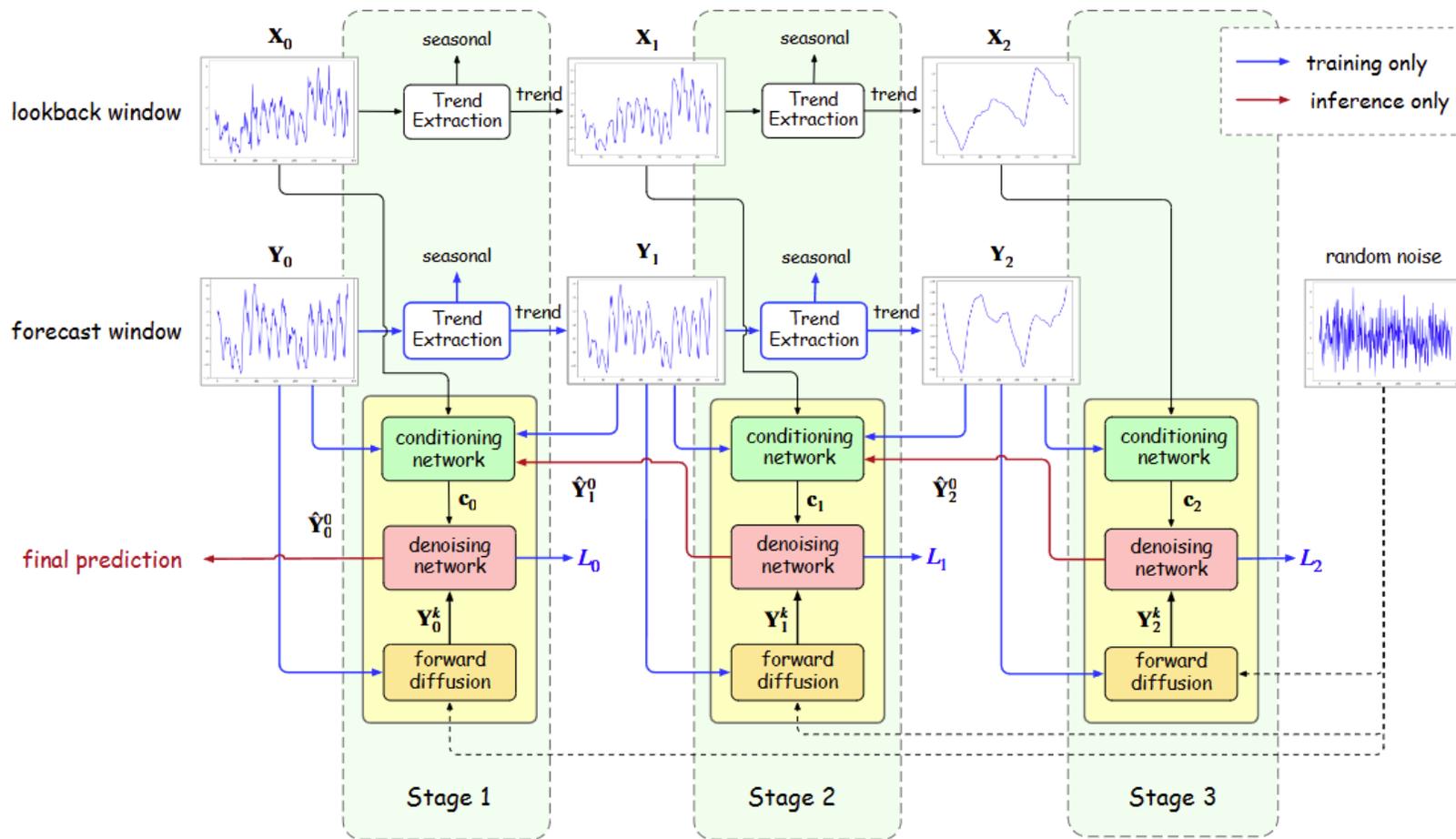
Multi-Resolution

- 시계열은 여러 scale의 패턴을 보이므로, 그, multi-resolution 구조를 활용
- 서로 다른 샘플링 scale의 시계열은 본질적으로 다른 특성을 보임
- 미세한 scale에서는 상세한 패턴을 보여주고, 거시적 scale에서는 전체적인 변동을 보여줌



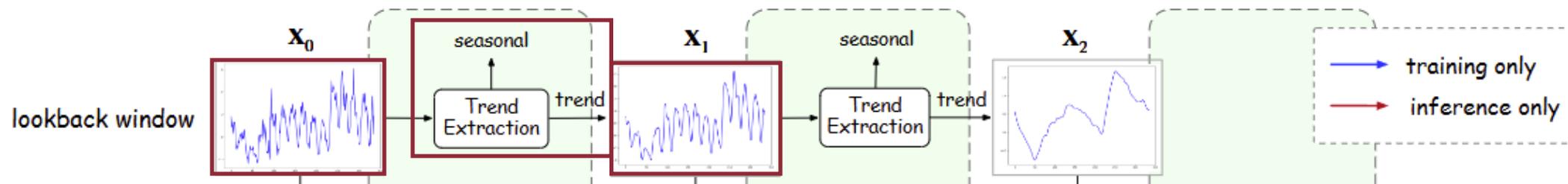


02 방법론

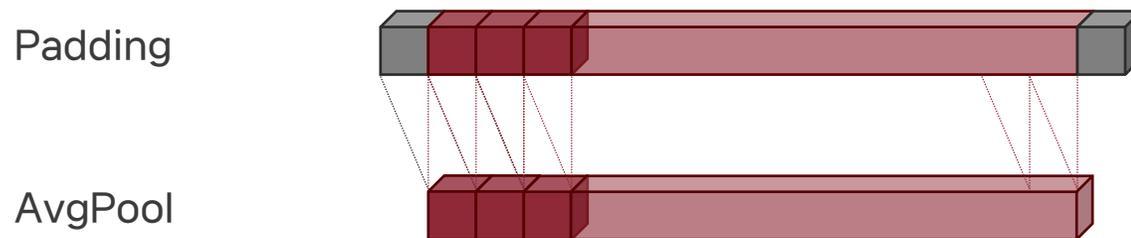




02 Multi-Resolution

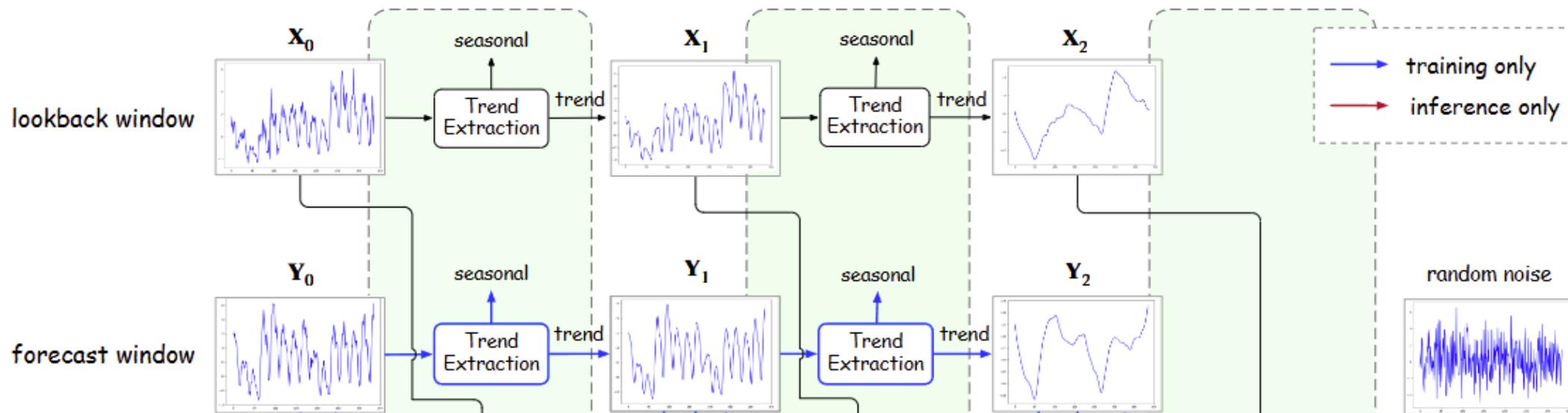


- ① 과거 데이터를 trend와 seasonal로 분해
- ② Trend 데이터를 coarser하게 만들





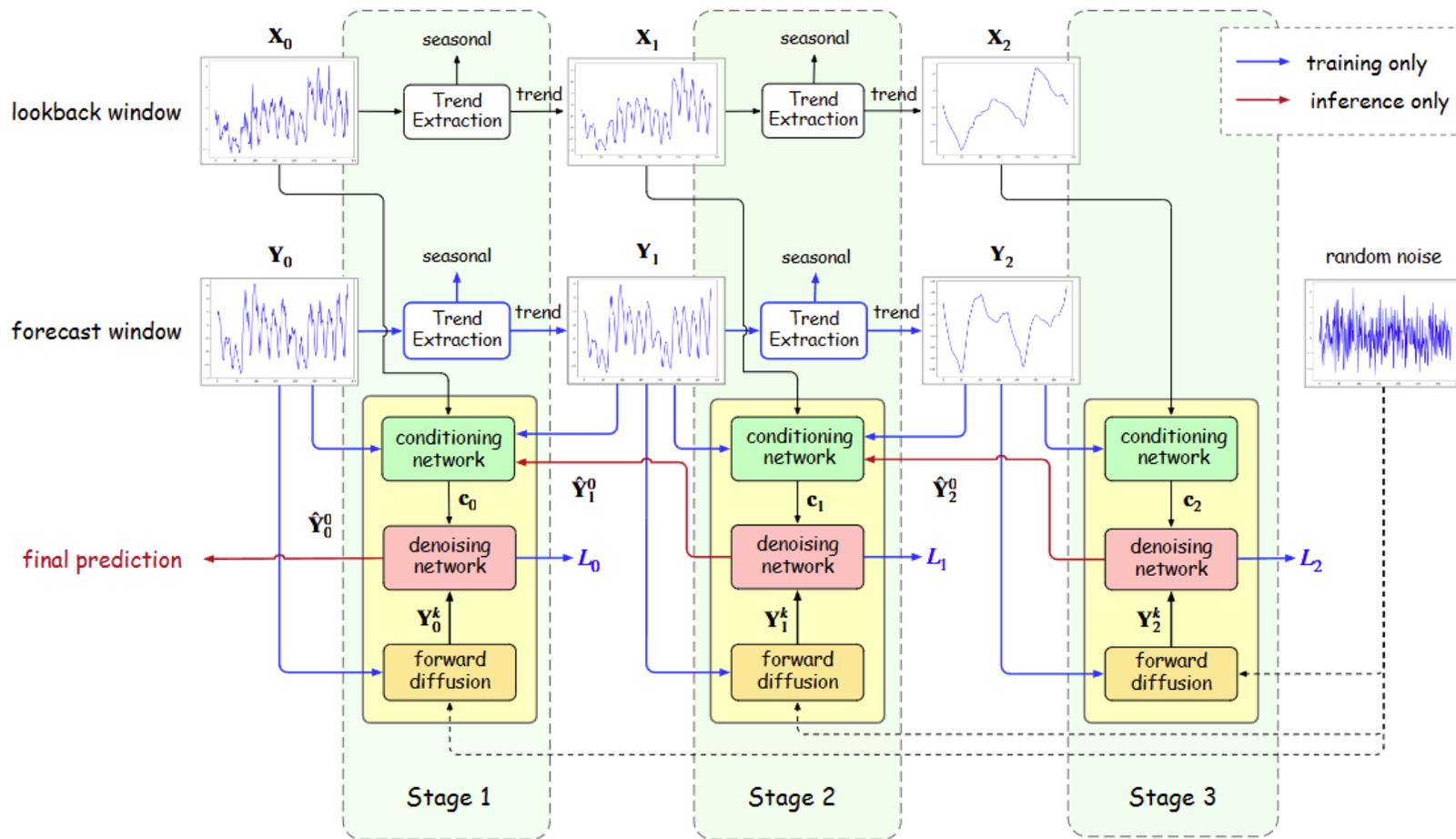
02 Multi-Resolution



- ① 과거 데이터를 trend와 seasonal로 분해
- ② Trend 데이터를 coarser하게 만듦
- ③ 예측하고자 하는 미래 데이터 역시 multi-resolution 형태의 데이터로 만듦

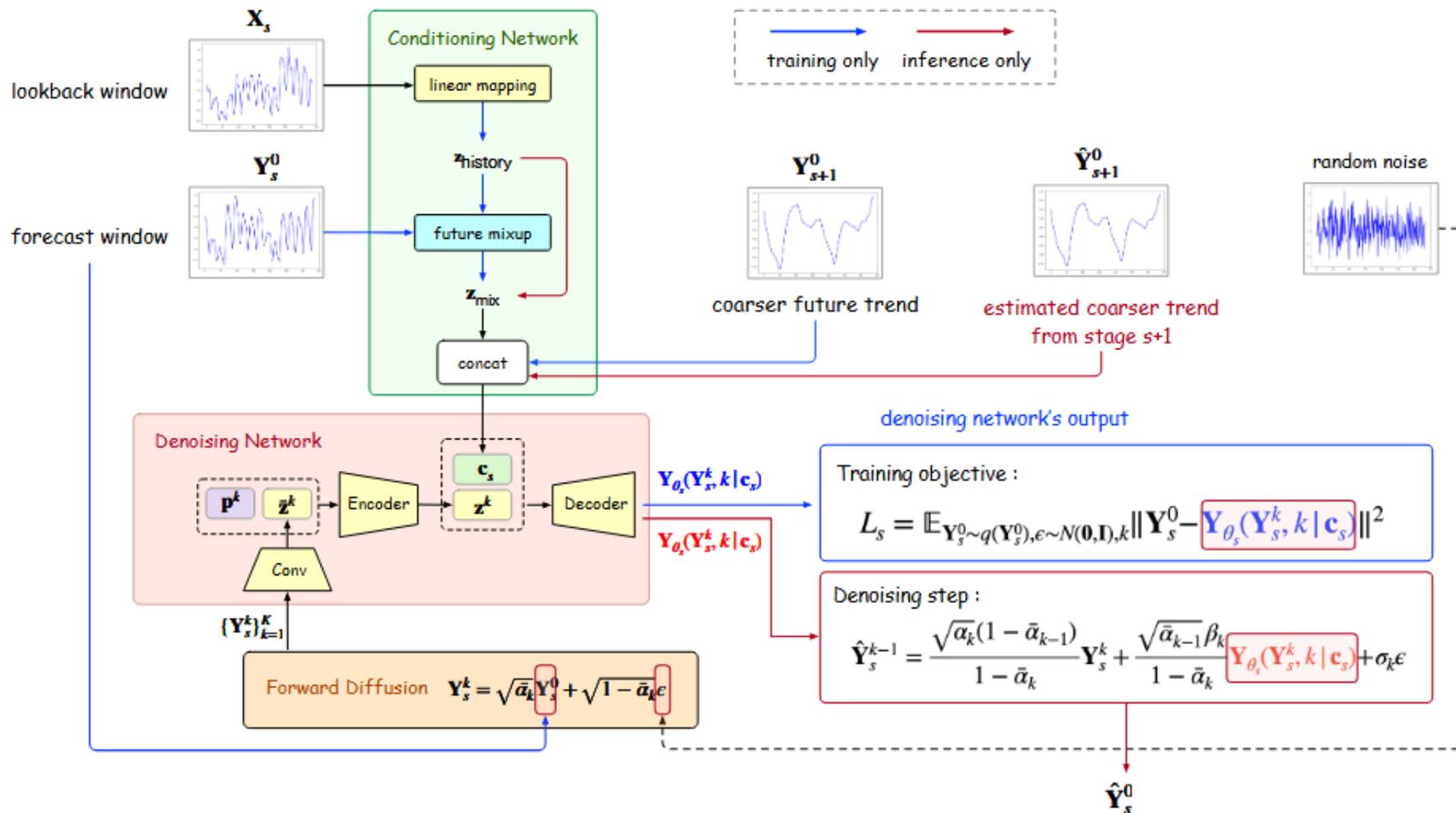


02 Multi-Resolution



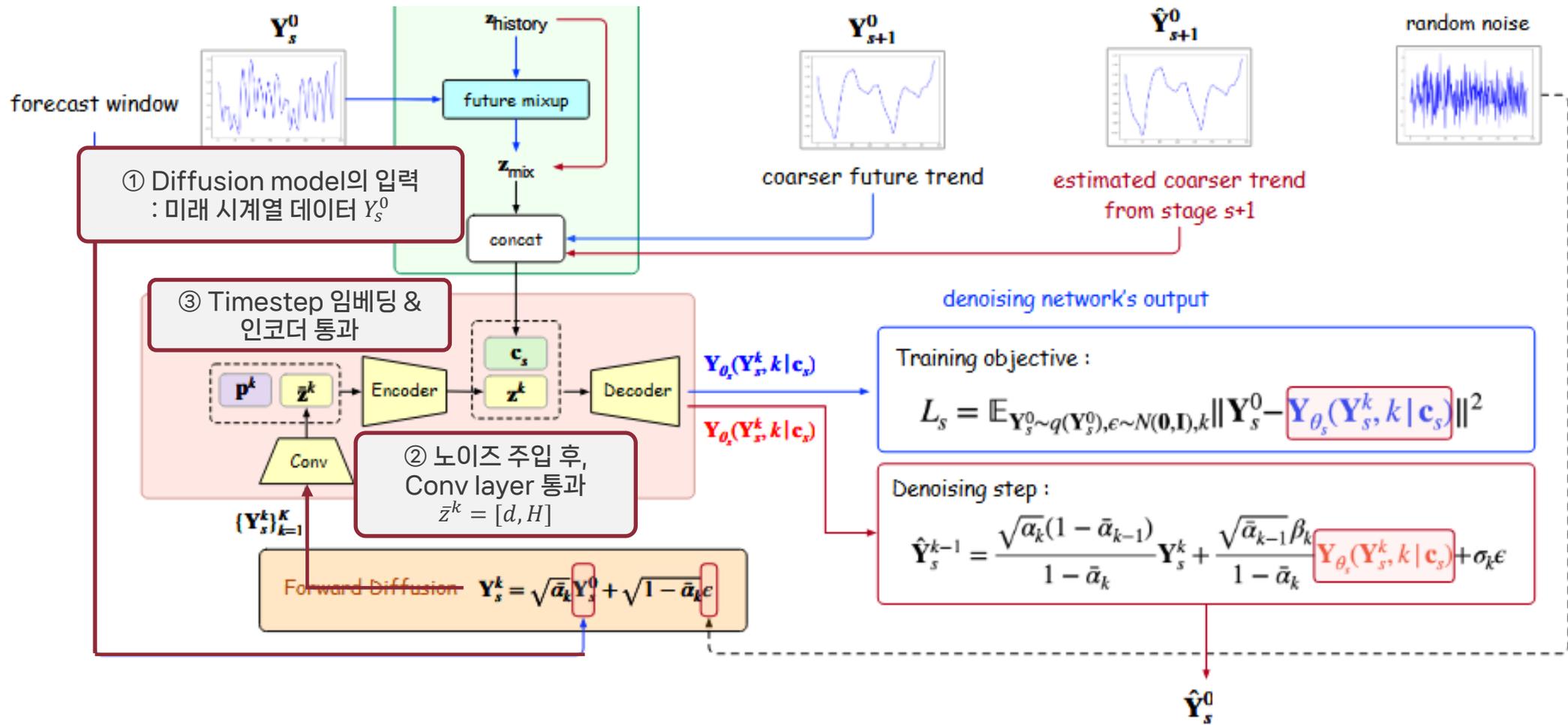


02 Denoising & Condition Network



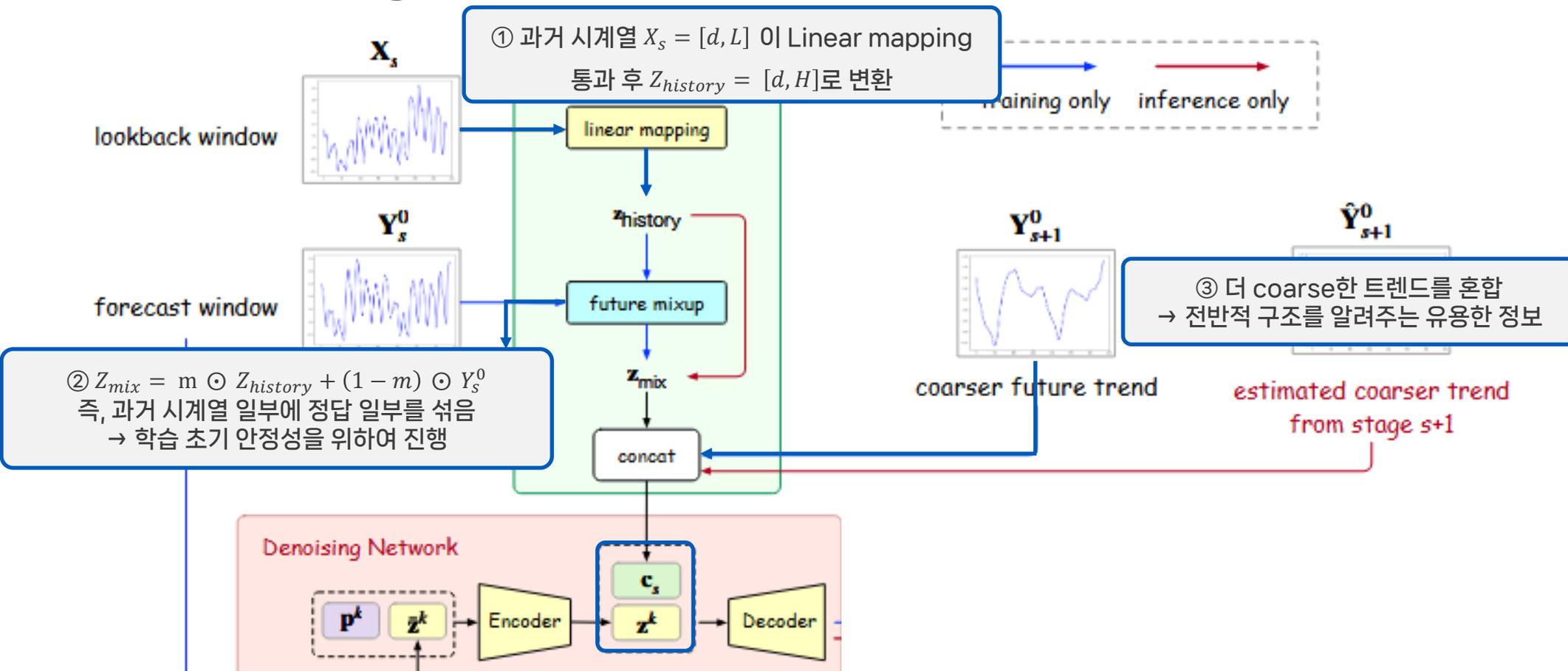


02 Denoising & Condition Network



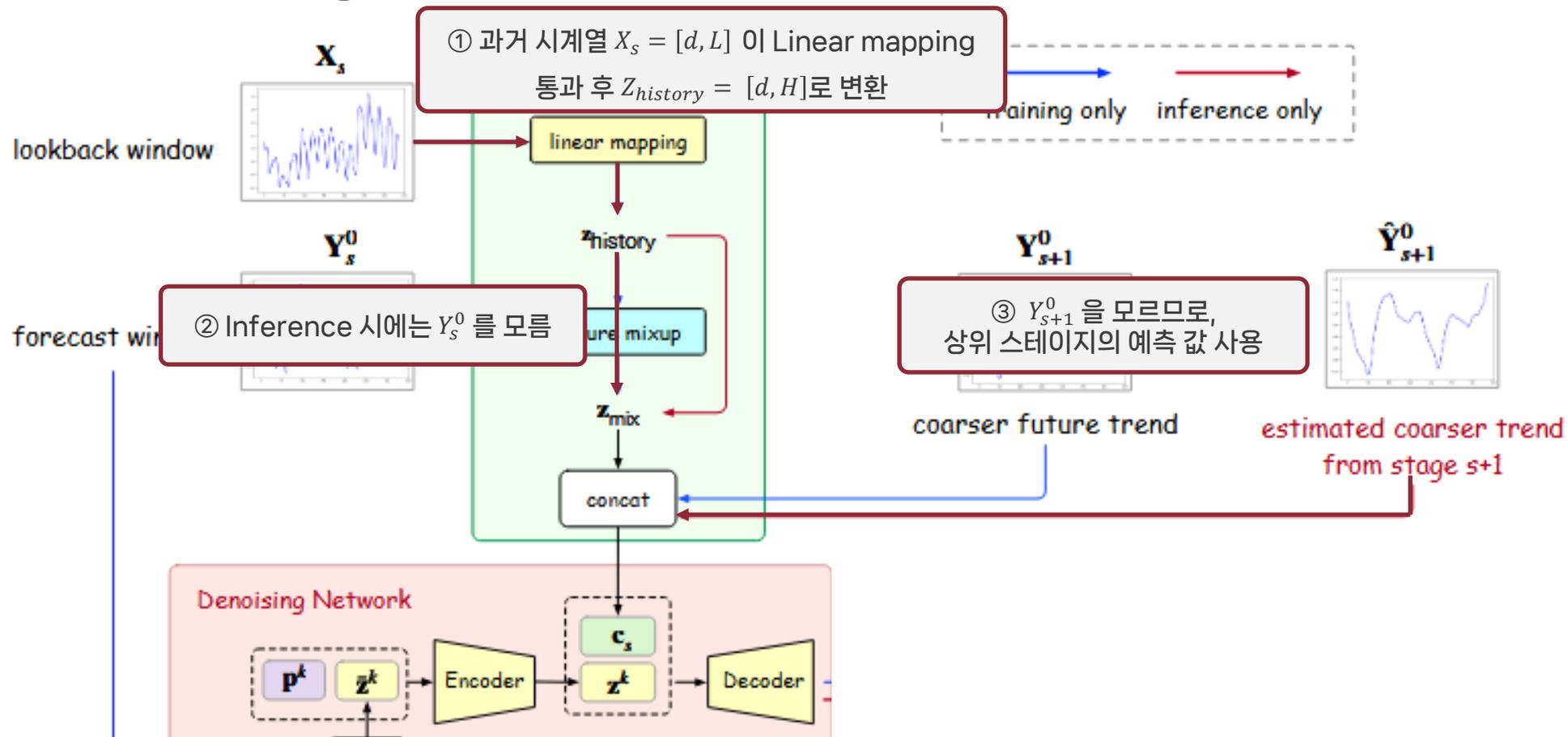


02 Denoising & Condition Network



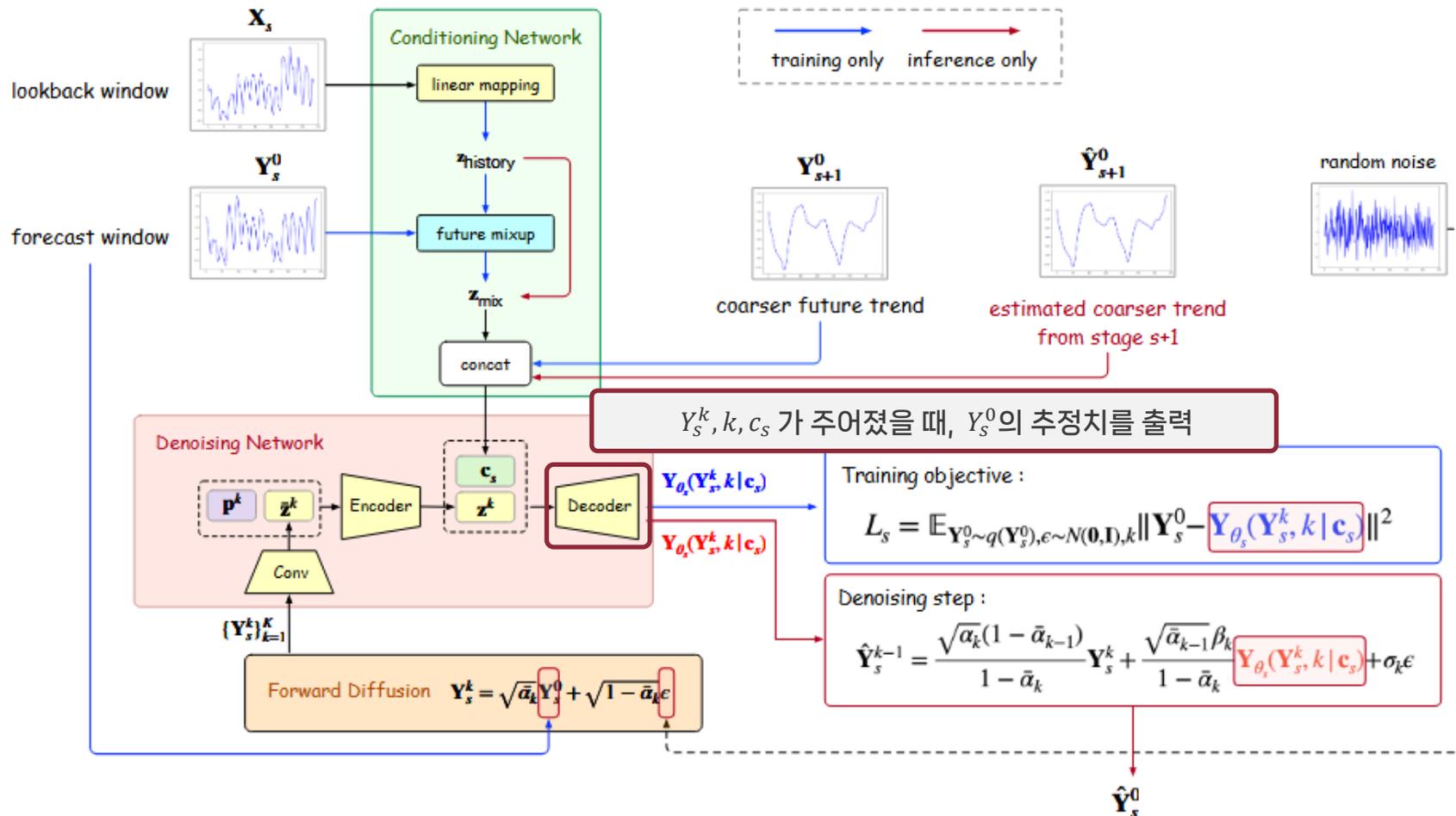


02 Denoising & Condition Network





02 Denoising & Condition Network





02 Conclusion

Results

- 다변량 시계열 작업에서의 MSE 값

Method	NorPool	Caiso	Traffic	Electricity	Weather	Exchange	ETTh1	ETTm1	Wind	Rank
mr-Diff	<u>0.645</u> ₍₂₎	0.127 ₍₃₎	0.474 ₍₆₎	0.155 ₍₃₎	0.296 ₍₁₎	0.016 ₍₁₎	0.411 ₍₃₎	<u>0.340</u> ₍₂₎	0.881 ₍₁₎	2.4
TimeDiff	0.665 ₍₄₎	0.136 ₍₆₎	0.564 ₍₇₎	0.193 ₍₅₎	<u>0.311</u> ₍₂₎	0.018 ₍₆₎	0.407 ₍₁₎	0.336 ₍₁₎	<u>0.896</u> ₍₂₎	3.8
TimeGrad	1.152 ₍₂₀₎	0.258 ₍₁₉₎	1.745 ₍₂₂₎	0.736 ₍₂₁₎	<u>0.392</u> ₍₁₄₎	0.079 ₍₂₀₎	0.993 ₍₂₂₎	0.874 ₍₂₁₎	1.209 ₍₂₁₎	20.0
CSDI	1.011 ₍₁₉₎	0.253 ₍₁₈₎	-	-	<u>0.356</u> ₍₉₎	0.077 ₍₁₉₎	0.497 ₍₇₎	0.529 ₍₁₇₎	1.066 ₍₉₎	14.0
SSSD	0.872 ₍₁₂₎	0.195 ₍₁₀₎	0.642 ₍₁₁₎	0.255 ₍₁₂₎	0.349 ₍₈₎	0.061 ₍₁₆₎	0.726 ₍₁₈₎	0.464 ₍₁₃₎	1.188 ₍₁₉₎	13.2
D ³ VAE	0.745 ₍₉₎	0.241 ₍₁₇₎	0.928 ₍₁₇₎	0.286 ₍₁₅₎	0.375 ₍₁₁₎	0.200 ₍₂₂₎	0.504 ₍₉₎	0.362 ₍₈₎	1.118 ₍₁₅₎	13.7
CPF	1.613 ₍₂₃₎	0.383 ₍₂₁₎	1.625 ₍₂₁₎	0.793 ₍₂₂₎	1.390 ₍₂₃₎	0.016 ₍₁₎	0.730 ₍₁₉₎	0.482 ₍₁₅₎	1.140 ₍₁₇₎	18.0
PSA-GAN	1.501 ₍₂₂₎	0.510 ₍₂₃₎	1.614 ₍₂₀₎	0.535 ₍₂₀₎	1.220 ₍₂₁₎	0.018 ₍₆₎	0.623 ₍₁₇₎	0.537 ₍₁₈₎	1.127 ₍₁₆₎	18.1
N-Hits	0.716 ₍₇₎	0.131 ₍₄₎	<u>0.386</u> ₍₂₎	<u>0.152</u> ₍₂₎	0.323 ₍₄₎	<u>0.017</u> ₍₅₎	0.498 ₍₈₎	0.353 ₍₆₎	1.033 ₍₆₎	4.9
FiLM	0.723 ₍₈₎	0.179 ₍₈₎	0.628 ₍₁₀₎	0.210 ₍₈₎	0.327 ₍₅₎	0.016 ₍₁₎	0.426 ₍₅₎	0.347 ₍₄₎	0.984 ₍₄₎	5.9
Depts	0.662 ₍₃₎	<u>0.106</u> ₍₂₎	1.019 ₍₁₉₎	0.319 ₍₁₇₎	0.761 ₍₁₉₎	0.020 ₍₉₎	0.579 ₍₁₃₎	0.380 ₍₁₀₎	1.082 ₍₁₂₎	11.6
NBeats	0.832 ₍₁₀₎	0.141 ₍₇₎	0.373 ₍₁₎	0.269 ₍₁₃₎	1.344 ₍₂₂₎	0.016 ₍₁₎	0.586 ₍₁₅₎	0.391 ₍₁₁₎	1.069 ₍₁₀₎	10.0
Scaleformer	0.983 ₍₁₅₎	0.207 ₍₁₃₎	0.618 ₍₉₎	0.195 ₍₆₎	0.462 ₍₁₆₎	0.036 ₍₁₂₎	0.613 ₍₁₆₎	0.481 ₍₁₄₎	1.359 ₍₂₂₎	13.7
PatchTST	0.851 ₍₁₁₎	0.193 ₍₉₎	0.831 ₍₁₆₎	0.225 ₍₁₀₎	0.782 ₍₂₀₎	0.047 ₍₁₄₎	0.526 ₍₁₁₎	0.372 ₍₉₎	1.070 ₍₁₁₎	12.3
FedFormer	0.873 ₍₁₃₎	0.205 ₍₁₁₎	0.591 ₍₈₎	0.238 ₍₁₁₎	0.342 ₍₇₎	0.133 ₍₂₁₎	0.541 ₍₁₂₎	0.426 ₍₁₂₎	1.113 ₍₁₄₎	12.1
Autoformer	0.940 ₍₁₄₎	0.226 ₍₁₅₎	0.688 ₍₁₅₎	0.201 ₍₇₎	0.360 ₍₁₀₎	0.056 ₍₁₅₎	0.516 ₍₁₀₎	0.565 ₍₁₉₎	1.083 ₍₁₃₎	13.1
Pyraformer	1.008 ₍₁₈₎	0.273 ₍₂₀₎	0.659 ₍₁₂₎	0.273 ₍₁₄₎	0.394 ₍₁₅₎	0.032 ₍₁₁₎	0.579 ₍₁₃₎	0.493 ₍₁₆₎	1.061 ₍₈₎	14.1
Informer	0.985 ₍₁₆₎	0.231 ₍₁₆₎	0.664 ₍₁₃₎	0.298 ₍₁₆₎	0.385 ₍₁₂₎	0.073 ₍₁₈₎	0.775 ₍₂₁₎	0.673 ₍₂₀₎	1.168 ₍₁₈₎	16.7
Transformer	1.005 ₍₁₇₎	0.206 ₍₁₂₎	0.671 ₍₁₄₎	0.328 ₍₁₈₎	0.388 ₍₁₃₎	0.062 ₍₁₇₎	0.759 ₍₂₀₎	0.992 ₍₂₂₎	1.201 ₍₂₀₎	17.0
SCINet	0.613 ₍₁₎	0.095 ₍₁₎	0.434 ₍₅₎	0.171 ₍₄₎	0.329 ₍₆₎	0.036 ₍₁₂₎	0.465 ₍₆₎	0.359 ₍₇₎	1.055 ₍₇₎	5.4
NLinear	0.707 ₍₆₎	0.135 ₍₅₎	0.430 ₍₄₎	0.147 ₍₁₎	0.313 ₍₃₎	0.019 ₍₈₎	<u>0.410</u> ₍₂₎	0.349 ₍₅₎	0.989 ₍₅₎	4.3
DLinear	0.670 ₍₅₎	0.461 ₍₂₂₎	0.389 ₍₃₎	0.215 ₍₉₎	0.488 ₍₁₇₎	0.022 ₍₁₀₎	0.415 ₍₄₎	<u>0.345</u> ₍₃₎	0.899 ₍₃₎	8.4
LSTMa	1.481 ₍₂₁₎	0.217 ₍₁₄₎	0.966 ₍₁₈₎	0.414 ₍₁₉₎	0.662 ₍₁₈₎	0.403 ₍₂₃₎	1.149 ₍₂₃₎	1.030 ₍₂₃₎	1.464 ₍₂₃₎	20.2



02 Conclusion

Results

- 다변량 시계열 작업에서의 MSE 값
- 트렌드 패턴이 복잡한 데이터에서 특히 좋은 성능
- 기존 diffusion 기반 시계열 모델들을 모두 능가하며, 그 이외의 다양한 고급 시계열 예측 모델들과 비교해도 더 나은 성능

Conclusion

- Diffusion 기반 시계열 모델에서 Trend-Seasonal 분해 기반 multi resolution 분석을 최초로 통합
- 더 coarse한 신호를 먼저 생성하고 점진적으로 디테일을 추가



D3U: Diffusion-based Decoupled Deterministic and Uncertain framework for probabilistic multivariate time series forecasting (ICLR, 2025 poster)

- Decoupled deterministic and uncertain
- 즉, 시계열 데이터에서 deterministic 부분과 uncertain 부분을 분리하여 예측하는 프레임워크

Diffusion-based Decoupled Deterministic and Uncertain Framework for Probabilistic Multivariate Time Series Forecasting



Qi Li, Zhenyu Zhang, Lei Yao, Zhaoxia Li, Tianyi Zhong, Yong Zhang

Published: 23 Jan 2025, Last Modified: 05 Mar 2025 ICLR 2025 Poster Everyone Revisions BibTeX CC BY 4.0

Keywords: long-term time series forecasting, deep learning, diffusion model

Abstract:

Diffusion-based denoising models have demonstrated impressive performance in probabilistic forecasting for multivariate time series (MTS). Nonetheless, existing approaches often model the entire data distribution, neglecting the variability in uncertainty across different components of the time series. This paper introduces a Diffusion-based Decoupled Deterministic and Uncertain (D^3U) framework for probabilistic MTS forecasting. The framework integrates non-probabilistic forecasting with conditional diffusion generation, enabling both accurate point predictions and probabilistic forecasting. D^3U utilizes a point forecasting model to non-probabilistically model high-certainty components in the time series, generating embedded representations that are conditionally injected into a diffusion model. To better model high-uncertainty components, a patch-based denoising network (PatchDN) is designed in the conditional diffusion model. Designed as a plug-and-play framework, D^3U can be seamlessly integrated into existing point forecasting models to provide probabilistic forecasting capabilities. It can also be applied to other conditional diffusion methods that incorporate point forecasting models. Experiments on six real-world datasets demonstrate that our method achieves over a 20% improvement in both point and probabilistic forecasting performance in MTS long-term forecasting compared to state-of-the-art (SOTA) probabilistic forecasting methods. Additionally, extensive ablation studies further validate the effectiveness of the D^3U framework.

Primary Area: generative models

Code Of Ethics: I acknowledge that I and all co-authors of this work have read and commit to adhering to the ICLR Code of Ethics.

Submission Guidelines: I certify that this submission complies with the submission instructions as described on <https://iclr.cc/Conferences/2025/AuthorGuide>.

Anonymous Url: I certify that there is no URL (e.g., github page) that could be used to find authors' identity.

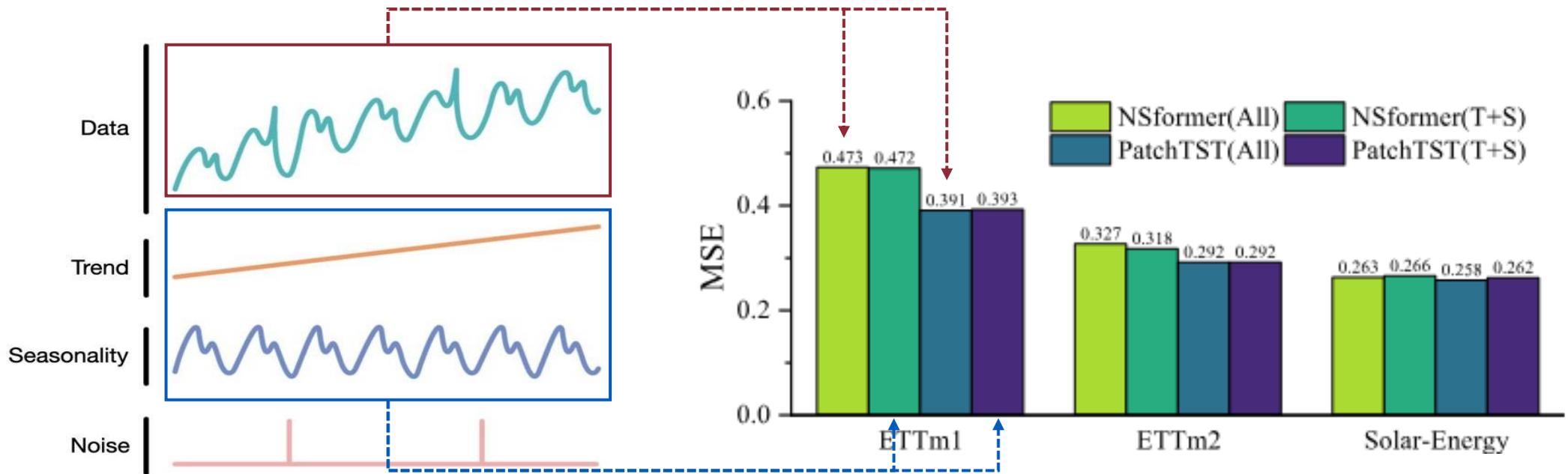
No Acknowledgement Section: I certify that there is no acknowledgement section in this submission for double blind review.

Submission Number: 7497

03 사례 연구

점 예측(point forecasting) 사례 연구

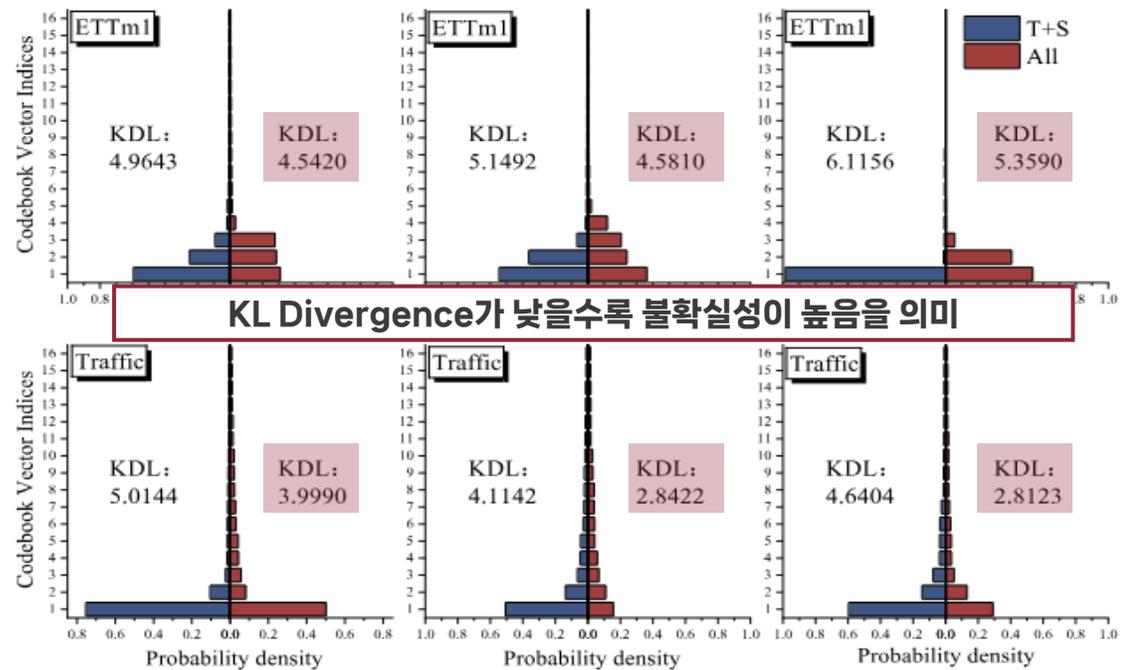
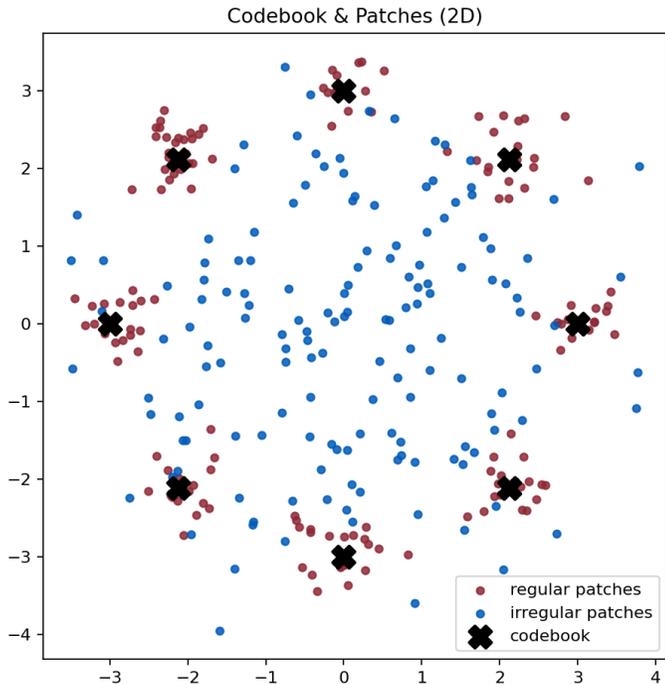
- 입력 시계열을 Trend(추세), Seasonality(계절), Residual(잔차) 성분으로 분해
- 잔차 성분을 제거했을 때, 모델이 성능이 저하되지 않거나 오히려 향상 → 잔차 성분의 모델링이 효과적이지 못함



03 사례 연구

벡터 양자화 (vector quantization) 모델 사례 연구

- 벡터 양자화는 입력 패치를 codebook vector 중 하나로 매핑
- 규칙적인 데이터는 특정 codebook vector에 집중되나, 불규칙 데이터는 고르게 분산되어 균등 분포에 가까움

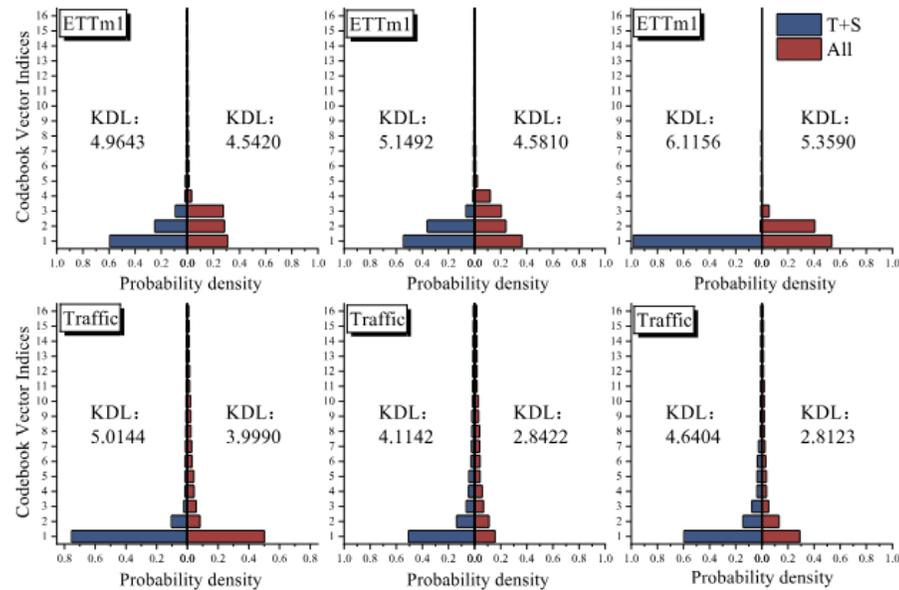
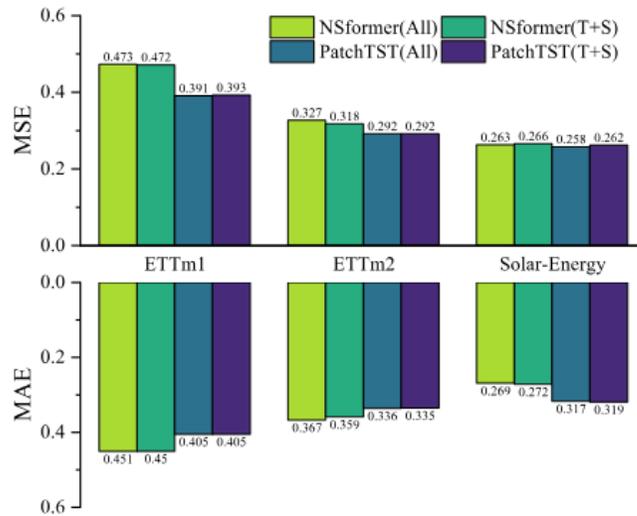


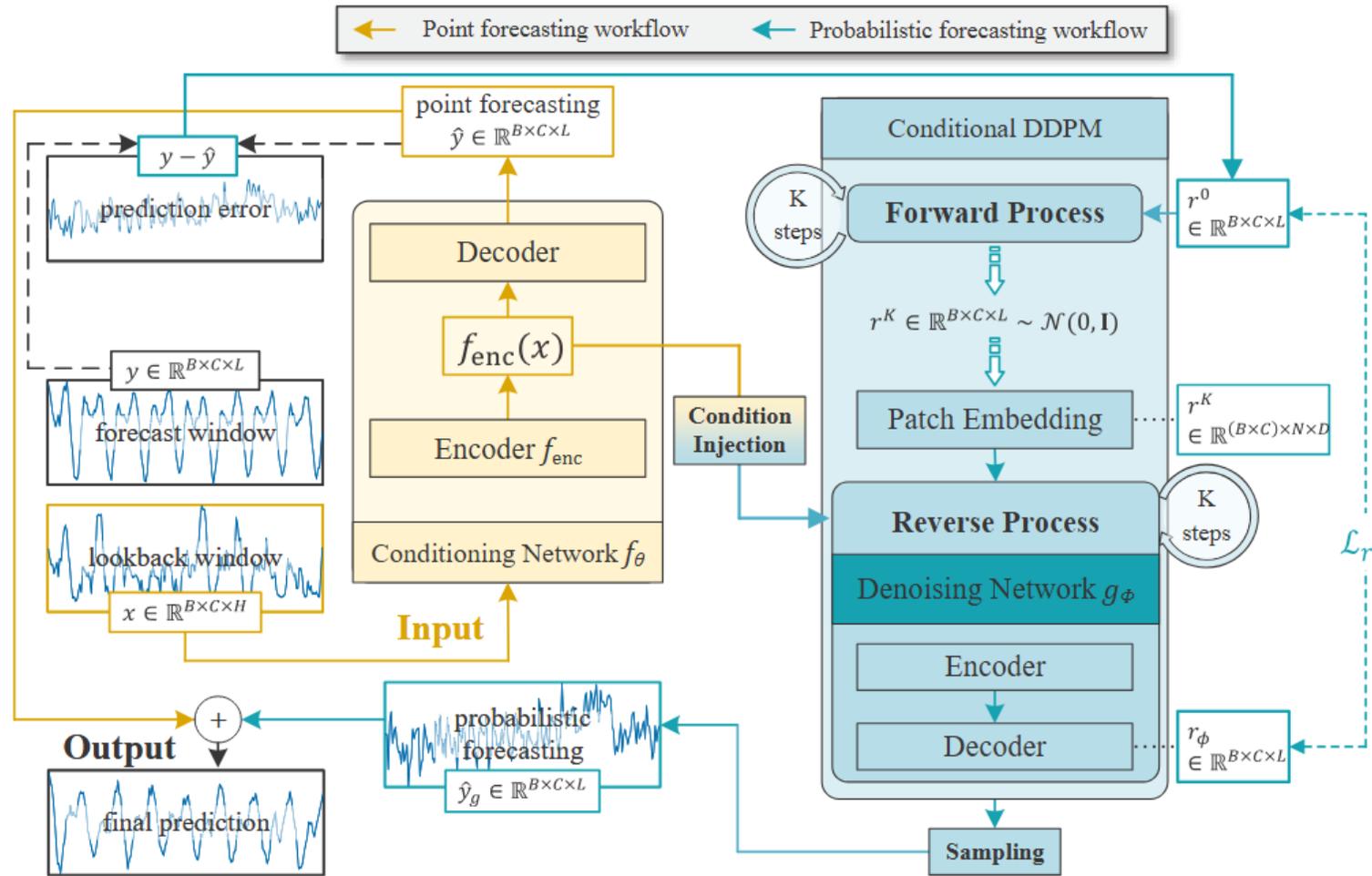
03 사례 연구

사례 연구 결과

- 기존 모델이 잔차 성분을 덜 효과적으로 모델링
- 잔차 성분은 추세, 계절 성분보다 더 많은 불확실성을 포함

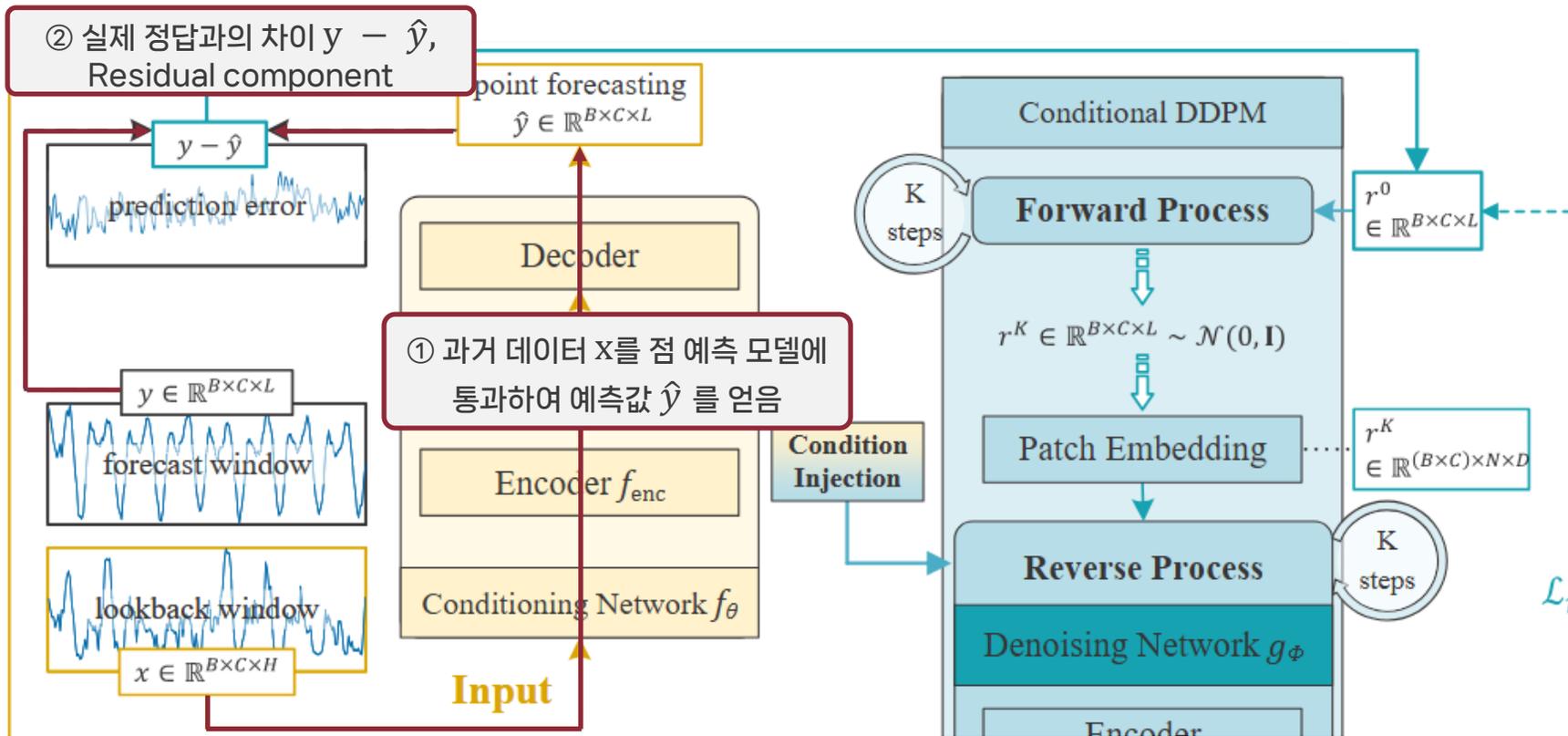
→ 확실성이 높은 성분은 점 예측 모델을 활용하고, 확실성이 낮은 잔차는 diffusion 모델로 처리하자





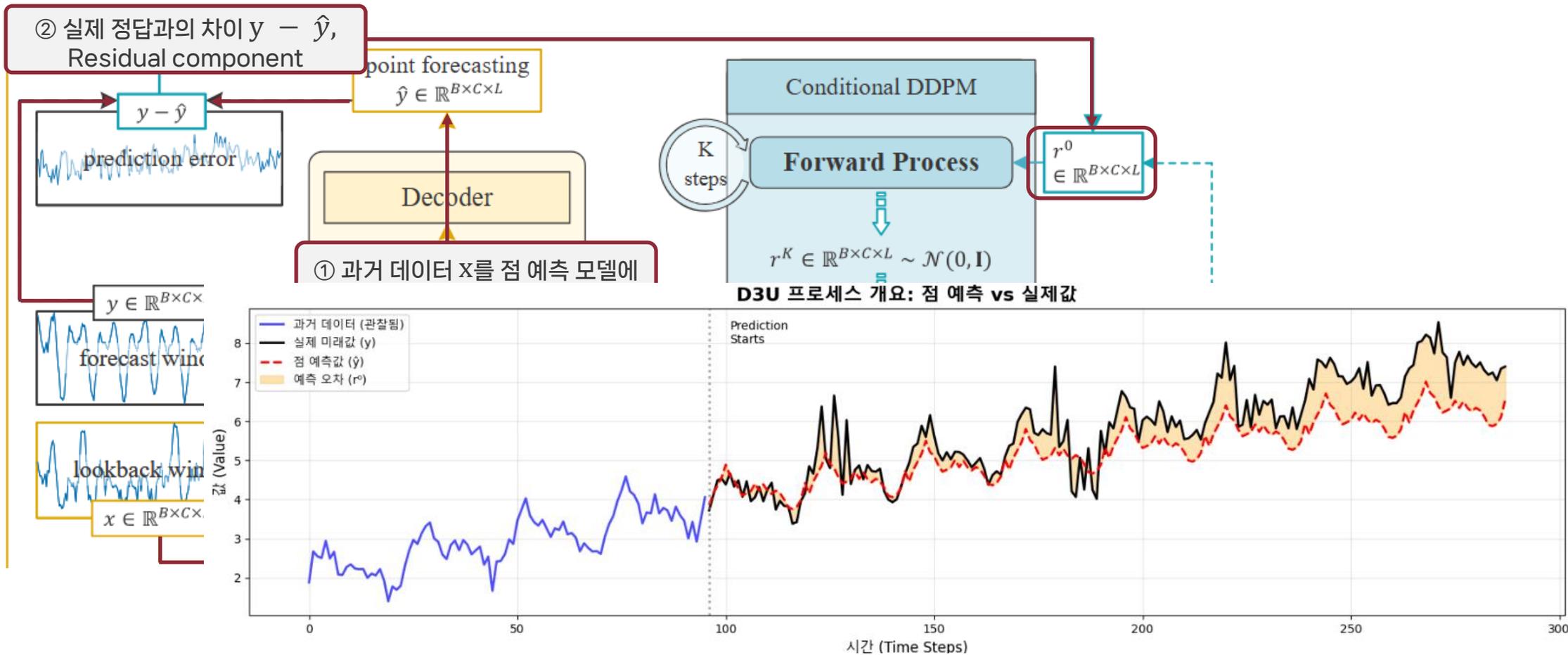


03 방법론 - forward process





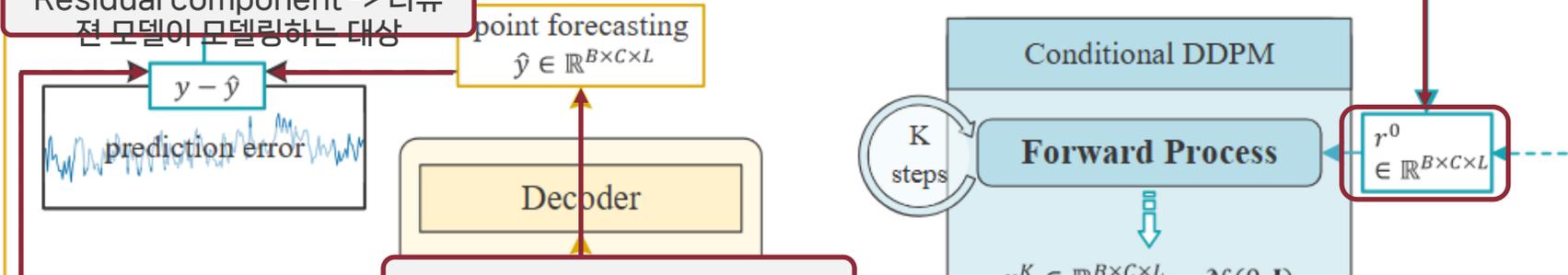
03 방법론 - forward process



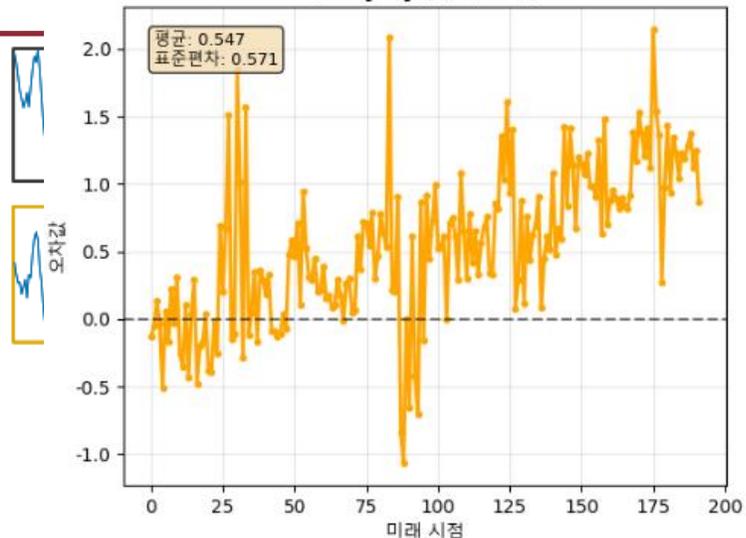


03 방법론 - forward process

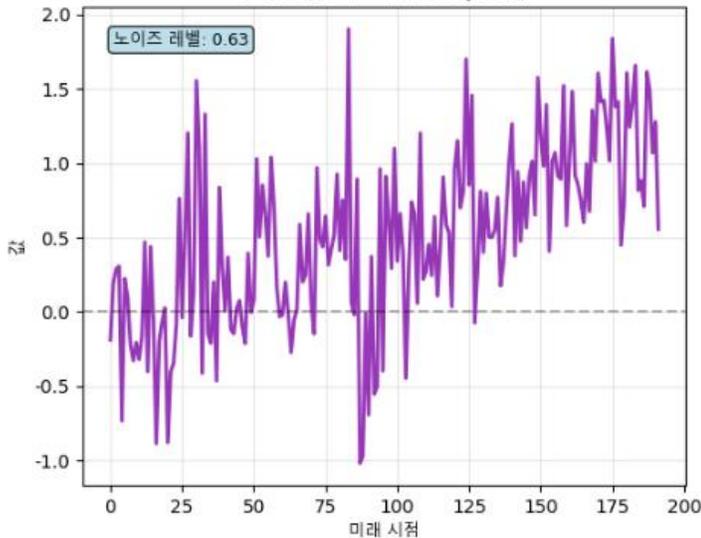
② 실제 정답과의 차이 $y - \hat{y}$, Residual component -> 디퓨전 모델이 모델링하는 대상



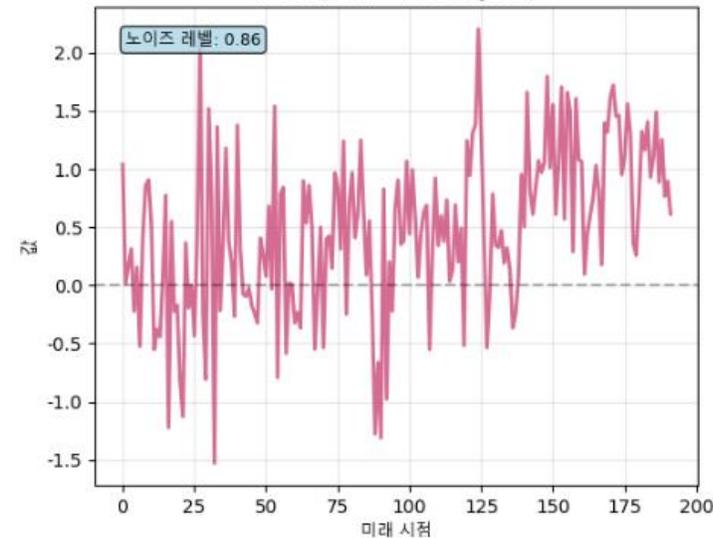
$r^0 = y - \hat{y}$ (예측 오차)



r^{25} (노이즈 주입 step 25)

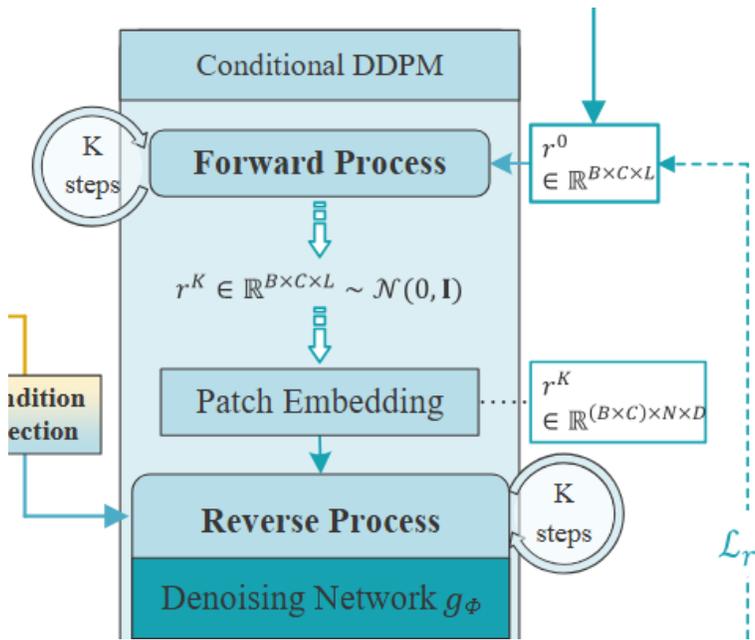


r^{50} (노이즈 주입 step 50)





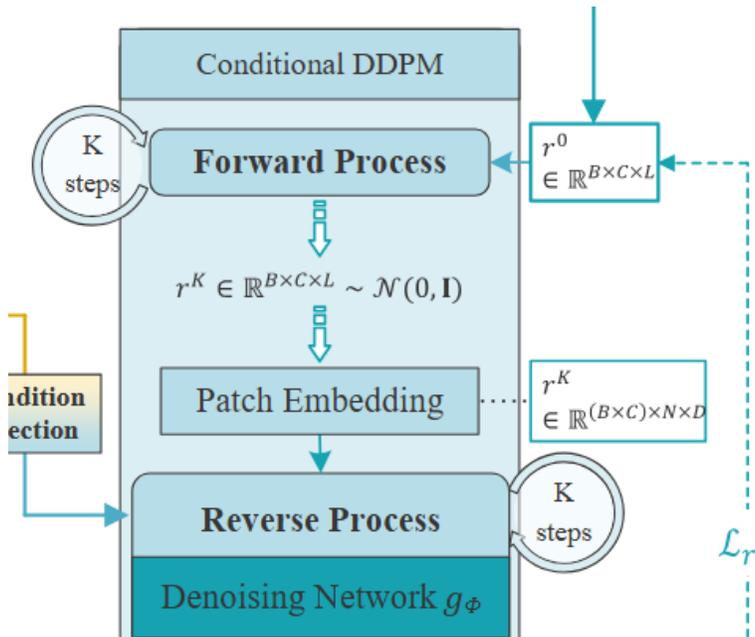
03 방법론 – patch embedding



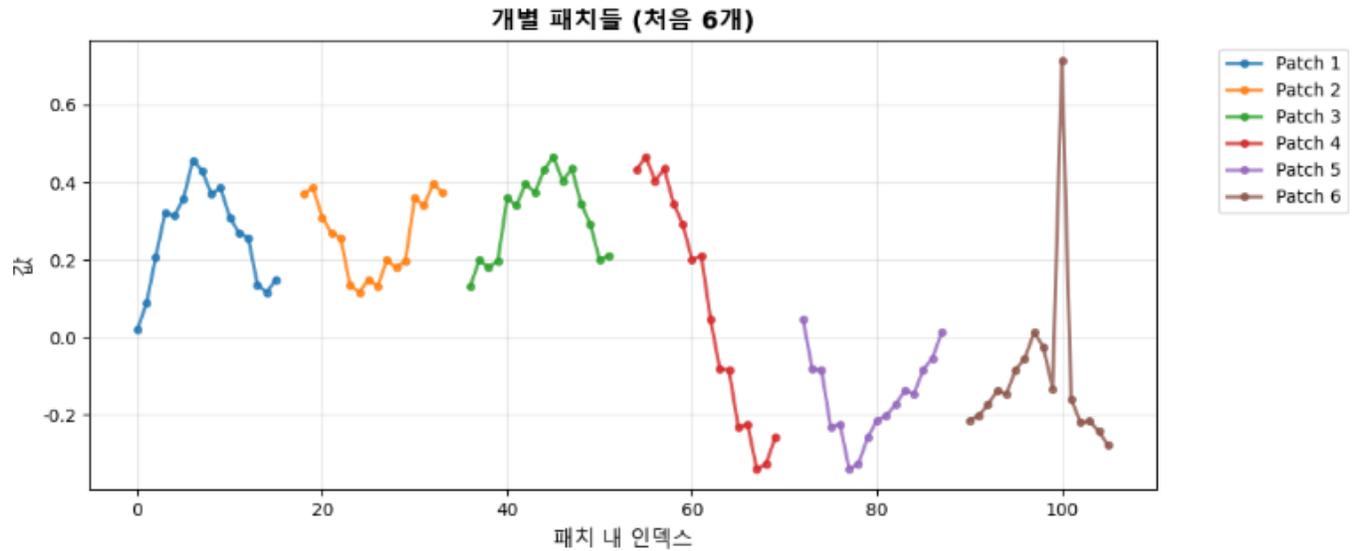
① r^k , 노이즈가 추가된 잔차 정보



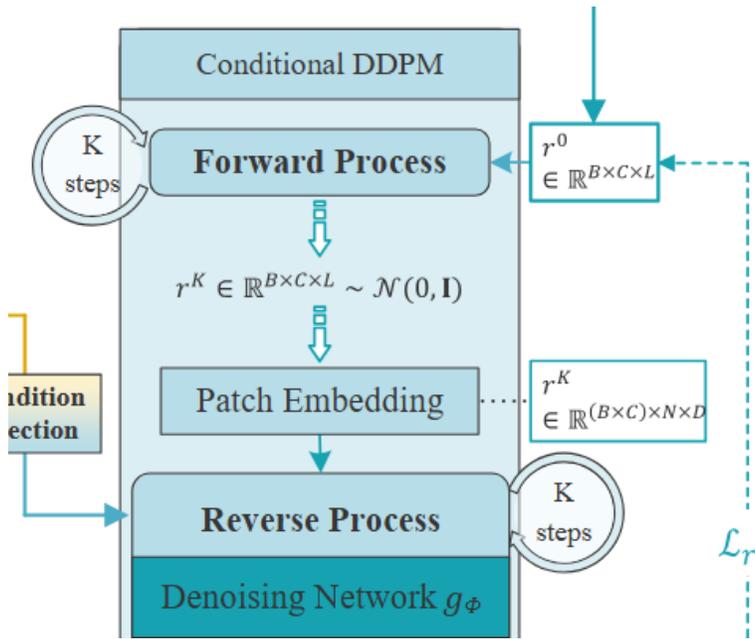
03 방법론 - patch embedding



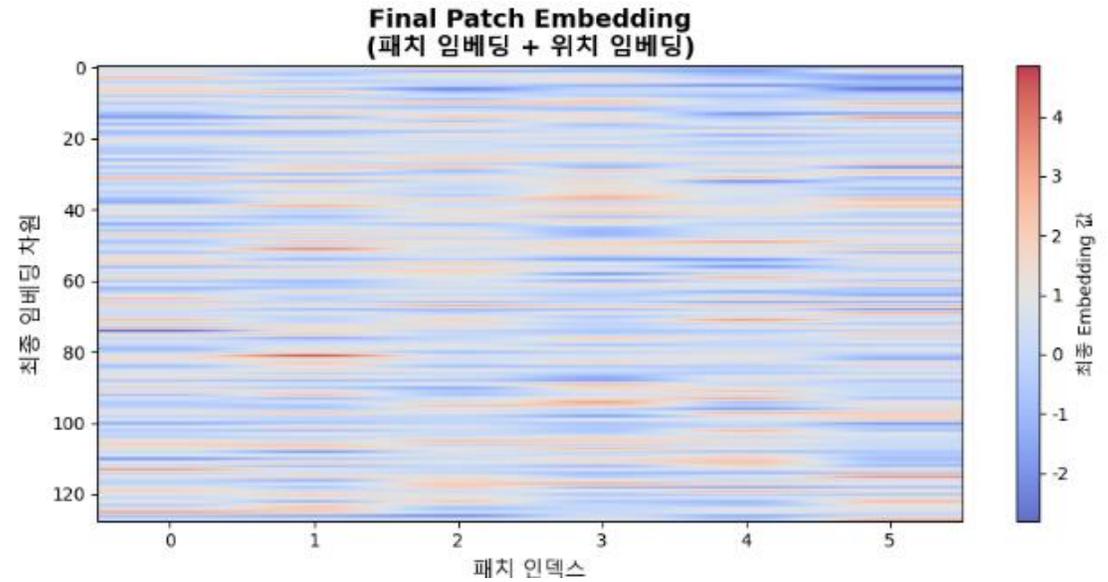
② 패치 단위로 분할



03 방법론 - patch embedding

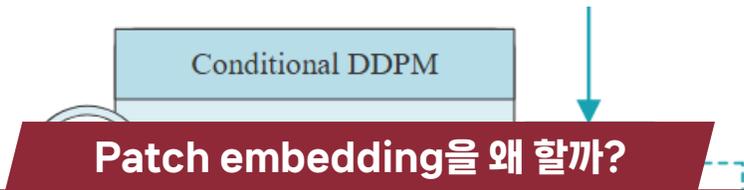


③ 패치를 선형 변환 (차원 확장) + positional embedding

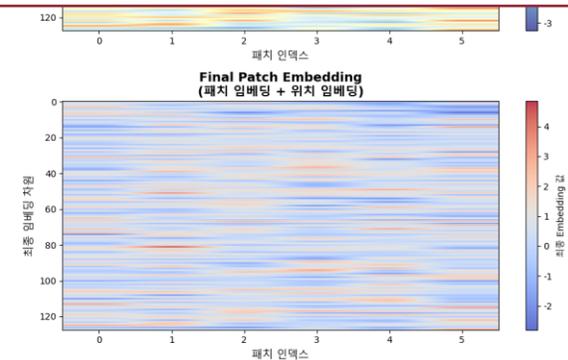
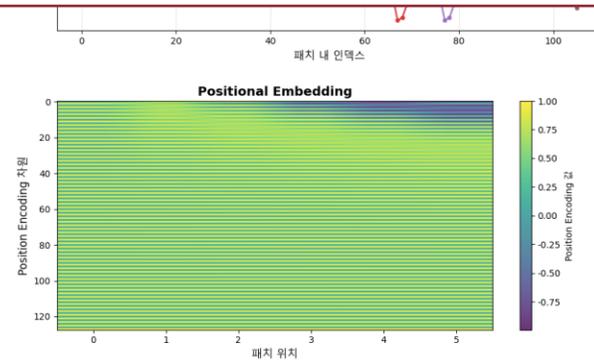
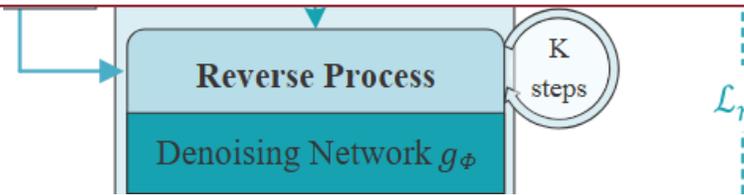




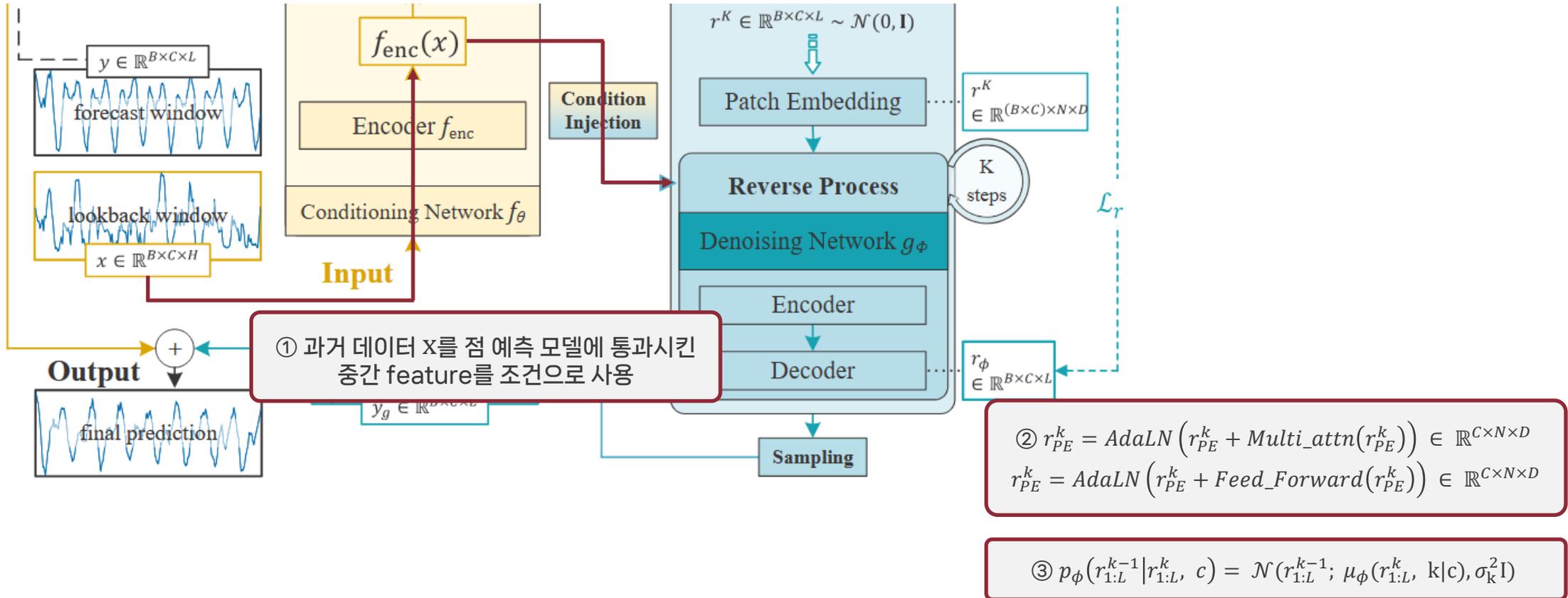
03 방법론 - patch embedding



1. 지역적 semantic 정보 효과적 포착
2. 각 시점에 대한 노이즈를 처리하면 불안정해지지만, 패치 내 노이즈 처리는 지역적 정보가 포함되어 더 안정적
3. Vision, Time Series(PatchTST, TimesNet) 등에서 우수성 입증



03 방법론 – reverse process



03 Conclusion

Results

- D3U는 대부분의 경우에서 다른 기준 모델들을 능가
- 다변량 시계열 데이터에서 deterministic 성분과 uncertain 성분을 분리하는 전략이 효과적임을 시사
- 조건 네트워크가 강력할수록 성능 향상에 중요한 역할

Model	Dataset	ETM1		ETM2		Weather		Solar-Energy		Electricity		Traffic	
	Method	MSE	MAE										
Point forecasting	NSformer(2022b)	0.440	0.430	0.277	0.343	0.226	0.270	0.266	<u>0.270</u>	0.191	0.295	0.653	0.360
	TimesNet(2023)	0.374	0.387	0.249	0.309	0.219	0.261	0.296	0.318	0.184	0.289	0.617	0.336
	DLinear(2023)	0.380	0.389	0.284	0.362	0.237	0.296	0.320	0.398	0.196	0.285	0.598	0.370
	PatchTST(2023)	<u>0.370</u>	0.390	0.251	0.312	0.223	<u>0.258</u>	0.259	0.321	0.205	0.307	<u>0.463</u>	0.311
	SparseVQ(2024)	0.363	0.380	<u>0.242</u>	0.302	0.225	<u>0.258</u>	0.256	0.286	0.182	<u>0.267</u>	0.480	0.300
	iTransformer(2024)	0.377	0.391	0.250	0.309	<u>0.221</u>	0.254	0.233	0.261	0.164	0.255	0.418	0.284
Probabilistic forecasting	TimeGrad(2021)	1.716	1.057	1.385	0.732	0.885	0.551	1.211	1.004	0.645	0.723	0.932	0.807
	CSDI(2021)	0.867	0.690	1.291	0.576	0.842	0.523	0.848	0.818	0.553	0.795	0.921	0.678
	TimeDiff(2023)	0.796	0.577	0.284	0.342	0.277	0.331	1.169	0.936	0.730	0.690	1.465	0.851
	TMDM(2024)	0.607	0.558	0.524	0.493	0.244	0.286	0.295	0.317	0.222	0.329	0.721	0.411
	ours	0.363	<u>0.386</u>	0.241	0.302	0.222	0.264	<u>0.237</u>	<u>0.270</u>	<u>0.179</u>	<u>0.267</u>	0.468	<u>0.299</u>

고맙습니다

참고 문헌

- [1] Su, C., Cai, Z., Tian, Y., Zheng, Z., & Song, Y. (2025). Diffusion Models for Time Series Forecasting: A Survey. arXiv preprint arXiv:2507.14507.
- [2] Shen, L., Chen, W., & Kwok, J. (2024). Multi-resolution diffusion models for time series forecasting. In The Twelfth International Conference on Learning Representations.
- [3] Li, Q., Zhang, Z., Yao, L., Li, Z., Zhong, T., & Zhang, Y. (2025). Diffusion-based decoupled deterministic and uncertain framework for probabilistic multivariate time series forecasting. In The Thirteenth International Conference on Learning Representations.