

3. Experiment

Data-Level Analysis

- 타 모델과 비교했을 때 성능이 유의하게 높았던 ETTm2, ETTh2 데이터셋에 대해서 왜 제안 모델의 성능이 더 좋았는지를 데이터 단계에서 분석을 수행하였음
- 사용한 지표는 피어슨 상관계수 평균 (Pearson Correlation, ρ), 자기상관함수 (Autocorrelation Function, ACF), 상관행렬 고유값의 정규화된 Shannon 엔트로피, 신호대잡음비 (Signal-to-Noise Ratio, SNR), 스파이크율 (Spike Rate) 의 5종임
- ρ 는 변수 간 선형 결합 강도를 측정
- ACF 는 한 시계열 Y_t 가 자기 자신을 시차 k 만큼 이동한 것과 얼마나 닮았는지 (클수록 일/주 패턴 뚜렷)
- Shannon 엔트로피는 높을 수록 채널 간 중복, 결합성이 높고 낮을 수록 채널 간 중복, 결합성이 낮음
- SNR 은 1일 이동평균을 신호로 보고, 잔차 (원 시계열-이동평균)을 잡음으로 설정하였음
- Spike Rate 은 채널 별 Z 스코어 $>3\sigma$ 비율의 평균 (전체 관측 중 각 채널의 평균적인 극단 값 ($|z|>3$) 비중을 의미)

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad R_k = \frac{C_k}{C_0}, C_k = \frac{1}{N} \sum_{t=1}^{N-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y}) \quad H = - \frac{\sum_{i=1}^C p_i \log p_i}{\log C} \quad SNR_{time} = \frac{Var(1\text{일 이동평균})}{Var(\text{잔차})}$$

왼쪽부터 피어슨 상관계수 (ρ), 자기상관함수 (ACF), Shannon 엔트로피, 신호대잡음비 (SNR)수식

3. Experiment

Data-Level Analysis

- 구체적인 지표 계산 ($|\rho|_{mean}$, ACF@day, SNR_{time})은 다음과 같이 진행됨
- $|\rho|_{mean}$: 변수쌍에 대한 피어슨 상관계수의 절대값 평균 (Data 평균)

$$|\rho|_{mean} = \frac{2}{C(C-1)} \sum_{i < j} |\rho(x_i, x_j)|$$

- ACF@day: 어제 같은 시각과 오늘 값이 얼마나 닮았는지를 1일 간격 지연에서 측정하여, 하루 주기성의 안정도를 요약한 값

$$\text{ACF}@day = \frac{1}{C} \sum_{c=1}^C \left\{ \frac{1}{2b+1} \sum_{k=P_{day}-b}^{P_{day}+b} \frac{\sum_{t=k}^{T-1} (x_t^{(c)} - \bar{x}^{(c)})(x_{t-k}^{(c)} - \bar{x}^{(c)})}{\sum_{t=0}^{T-1} (x_t^{(c)} - \bar{x}^{(c)})^2 + \varepsilon} \right\}$$

C: 시계열의 채널 (변수) 개수, ETT = 7, Weather = 21

$|\rho(x_i, x_j)|$: 피어슨 상관계수

합의 범위 $i < j$: 서로 다른 모든 채널 쌍 평균

$x_t^{(c)}$: 채널 c, 시점 t에서의 시계열 관측값

$\bar{x}^{(c)}$: 채널 c의 표본 평균

T: 상관을 계산하는 시계열 길이 (각 데이터셋의 Train 구간 길이)

P_{day} : 하루에 해당하는 지연 (Lag)의 스텝 수 (Weather: 144, ETTm: 96, ETTh: 24)

b: 1, 주변 $\pm b$ 개 지연을 평균하여 주변 분산을 완화함

ε : 수치 안정화를 위한 상수, 10^{-9}

3. Experiment

Data-Level Analysis

- 구체적인 지표 계산 ($|\rho|_{mean}$, ACF@day, SNR_{time})은 다음과 같이 진행됨
- SNR_{time} : 하루 단위로 변하는 추세 · 주기 (신호)의 세기가 그 주변의 빠른 흔들림 (잡음) 보다 얼마나 큰지를 시간 영역 분산 비율로 측정한 값

$$s_t^{(c)} = \frac{1}{P_{day}} \sum_{j=-(\frac{P_{day}-1}{2})}^{\frac{P_{day}-1}{2}} x_{t+j}^{(c)}, \quad r_t^{(c)} = x_t^{(c)} - s_t^{(c)}$$
$$SNR_{time}^{(c)} = \frac{\text{Var}(s^{(c)})}{\text{Var}(r^{(c)}) + \varepsilon}, \quad SNR_{time} = \frac{1}{C} \sum_{c=1}^C SNR_{time}^{(c)}$$

C: 시계열의 채널 (변수) 개수, ETT = 7, Weather = 21

$x_t^{(c)}$: 채널 c , 시점 t 에서의 시계열 관측값

$s^{(c)}$: 저주파 (느린 변화) 성분

$r^{(c)}$: 잔차 (고주파 · 불규칙) 성분

$\text{Var}(\cdot)$: 표본 분산

P_{day} : 하루에 해당하는 지연 (Lag)의 스텝 수 (Weather: 144, ETM: 96, ETh: 24)

ε : 수치 안정화를 위한 상수, 10^{-9}

3. Experiment

Data-Level Analysis

- ρ 가 ETTm2, h2 ≈ 0.324 로 다른 데이터셋보다 채널들이 서로 더 촘촘히 연결되어 있음
- Shannon Entropy 를 보았을 때 ETTm1, h1 보다 상대적으로 소수의 채널이 데이터를 주도함을 확인할 수 있음
- ACF 가 높아 일/주 단위 주기성이 매우 선명하다고 생각할 수 있음
- 스파이크율을 보았을 때 ETTm2, h2 가 1.70%로 이상치 . 급변 이벤트가 조금 더 많이 관측됨을 알 수 있었음
- SNR 이 ETTm1, h1 보다 약 3배 높은데, SNR 이 높다는 것은 데이터의 저주파 성분이 잔차보다 우세하다는 의미 (평탄 · 주기적인 특성을 갖는다는 의미)

Datasets	Resolutions	The number of variances	$ \rho _{mean}$	Shannon Entropy	ACF@day	Spike Rate	SNR_{time}
ETTh1	hourly	7	0.2221	0.7799	0.8021	0.0120	6.5134
ETTm1	15min	7	0.2243	0.7787	0.7991	0.0119	6.4908
ETTh2	hourly	7	0.3246	0.7533	0.8231	0.0170	22.8323
ETTm2	15min	7	0.3245	0.7537	0.8243	0.0170	22.7629
weather	10min	21	0.2956	0.6275	0.5904	0.0083	7.9551

데이터 단위 분석 결과

3. Experiment

Data-Level Analysis

- 이에 다음과 같은 6개 가설을 세워 데이터 특성에 따른 모델 이득을 검증해보고자 함
 - H1: 일 · 주 ACF 를 약화시키면 제안 모델의 성능 하락이 ETTh2, m2 에서 더 클 것
 - H2: 임의 채널을 섞거나 일부를 제거하는 등 채널 간 상관구조(공변구조)를 약화시키면 성능 하락이 ETTh2, m2 에서 더 크게 나타날 것
 - H3: Dynamic Depth-Wise Conv 생성 시 이를 고정하거나 정적 Depth-Wise 로 교체하면 성능 하락이 ETTm2, h2 에서 가장 클 것
 - H4: 고주파를 증폭해서 SNR 을 낮추면 성능 하락이 ETTh2, m2 에서 크게 나타날 것
 - H5: Co-Attention Block 을 단순 결합으로 교체하면 ETTh2, ETTm2 에서 성능 하락이 뚜렷할 것
 - H6: RevIN (Reversible Instance Normalization) 을 OFF 하면 장기 horizon 에서 성능 하락이 나타날 것

3. Experiment

Data-Level Analysis

- 각 가설을 통해 검증해보고자 하는 것은 다음과 같음
 - H1: 모델이 일·주 단위 주기와 정렬되어 작동할 때 성능 이득이 큰가?
 - H2: 상관구조의 강도/집중도가 커질수록 채널 혼합·스트림 공유의 이득이 커지는가?
 - H3: 조건 벡터로 Kernel 을 생성하는 Dynamic Depth-Wise 메커니즘이 주기 위상·진폭 변조에 실제로 기여하는가?
 - H4: 저주파 골격 (평탄·주기적인 특성)이 모델 예측의 주요한 단서인가?
 - H5: Conv 토큰과 GRU 문맥을 Attention 으로 결합하는 것이 단순 결합보다 실제 이득인가?
 - H6: RevIN 이 장기 예측에서 분포 드리프트 (Distribution Drift) 를 실제로 줄여 성능에 기여하는가?

3. Experiment

Data-Level Analysis 총론

- 각 실험에 대하여 가설 검증 및 데이터 특성에 따른 모델 이득을 검증한 바는 다음과 같음
- **H1:** ETTm2, ETTh2 는 일·주 주기의 ACF, SNR 이 더 선명, 제안 모델은 이런 주기의 정합을 강하게 활용하여 성능 이득을 만들어 내고, 주기성이 약한 ETTm1, ETTh1 에서는 상대적으로 이득이 작았음
- **H2:** ETTm2, ETTh2 는 채널 간 상관관계보다 “같은 시점에서 함께 등장” 하는 시간 정합에서 제안 모델로 인한 이득이 발생하는 것으로 확인되고, α Cross Stitch 로 인한 이득은 미약한 것으로 확인됨
- **H3:** ETTm2, ETTh2 는 일 주기의 정합이 Datasets 중 가장 강해 작은 위상·진폭 차이를 시점 별 Kernel 로 즉시 보정하는 Dynamic Depth-Wise 가 성능을 개선하는 것으로 해석됨
- **H4:** SNR_{time} 을 낮추면 RMSE 가 선형적으로 증가하여 제안 모델이 저주파 골격에 의존함이 확인되었고, 민감도는 ETTm2 > ETTm1, ETTh1 > ETTh2, Weather 는 채널 이질성으로 둔감하게 나타남
- **H5:** ETTm2, ETTh2 는 Day / Week 정합 지수 (DAI / WAI) > 0 으로 “어제 / 지난주 같은 시간” 의 정보가 유용
· 안정하고, 이 조건에서 Co-Attention 의 선택 결합이 Concat 대비 일관된 성능 이득으로 나타남
- **H6:** ETTm2, ETTh2 는 고(高) 드리프트 데이터이고, RevIN 은 예측에서 확실한 이득을 줌, ETTm2 는 드리프트 완충이 주효 ($\mu \cdot \sigma^2$ 개선), ETTh2 μ 정렬과 학습 안정화 효과가 주효함

3. Experiment

Data-Level Analysis: H1

- H1 실험에서는 다음의 두 조건으로 실험을 수행함
 - (i) 일 · 주 주기 길이만큼 Data 를 나눈 뒤, 각 구간의 일부 시점을 섞어 같은 시간대의 반복 패턴 (위상 정렬) 을 깨고 ACF 를 크게 줄이는 Shuffle 실험을 수행함
 - (ii) 해당 주기의 저주파 성분을 제거하는 차분 실험 (Seasonal Differencing, Sdiff) 을 수행함
- H1 실험 결과, (ii) 실험에서 일 · 주 ACF 를 약화시키면 ETTm2, h2 에서의 성능이 ETTm1, h1 에서보다 더 크게 하락하는 것이 관찰되어, 주기 정합이 모델의 핵심 이득원이라고 해석됨
- (i) 실험에서는 ETTm1, h1 의 하락폭이 더 큰 경우가 많았는데, 이는 (i) 실험에서 ETTm1, h1 Datasets 의 $|\Delta ACF|$ 감소폭이 더 컸기 때문이라고 생각됨 (Shuffle 실험에서 일 주기, 주 주기와 RMSE 간의 상관이 매우 높게 관측됨, 일 주기와의 상관 ≈ 0.970 , 주 주기와의 상관 ≈ 0.917)

3. Experiment

Data-Level Analysis: H1

Datasets	Baseline RMSE (mean ± std)	Shuffle day Δ RMSE %	Shuffle day ΔACF day	Shuffle week Δ RMSE %	Shuffle week ΔACF week	Sdiff day Δ RMSE %	Sdiff day ΔACF day	Sdiff week Δ RMSE %	Sdiff week ΔACF week
ETTh1	3.055 ± 0.008	+ 49.8 %	0.153	+ 49.0 %	0.141	+ 98.1 %	0.456	+ 27.7 %	0.301
ETTh2	3.597 ± 0.013	+ 7.6 %	0.025	+ 3.6 %	0.023	+ 118.5 %	0.618	+ 42.0 %	0.343
ETTm1	2.818 ± 0.007	+ 56.8 %	0.152	+ 58.2 %	0.141	+ 62.5 %	0.449	+ 24.4 %	0.297
ETTm2	2.958 ± 0.006	+ 10.3 %	0.028	+ 17.1 %	0.021	+ 223.6 %	0.614	+ 29.1 %	0.343
Weather	39.672 ± 0.131	+ 34.5 %	0.125	+ 34.8 %	0.116	+ 342.9 %	0.290	+ 38.0 %	0.089

Horizon = 96에서의 실험 결과

Datasets	Baseline RMSE (mean ± std)	Shuffle day Δ RMSE %	Shuffle day ΔACF day	Shuffle week Δ RMSE %	Shuffle week ΔACF week	Sdiff day Δ RMSE %	Sdiff day ΔACF day	Sdiff week Δ RMSE %	Sdiff week ΔACF week
ETTh1	3.185 ± 0.015	+ 45.1 %	0.153	+ 45.1 %	0.142	+ 86.5 %	0.455	+ 29.4 %	0.299
ETTh2	3.655 ± 0.012	+ 7.2 %	0.025	+ 4.4 %	0.023	+ 161.9 %	0.620	+ 35.8 %	0.343
ETTm1	3.079 ± 0.006	+ 48.1 %	0.152	+ 47.3 %	0.141	+ 53.3 %	0.449	+ 24.6 %	0.297
ETTm2	3.343 ± 0.022	+ 8.3 %	0.028	+ 8.4 %	0.021	+ 184.1 %	0.614	+ 29.3 %	0.353
Weather	40.824 ± 0.176	+ 33.0 %	0.125	+ 31.6 %	0.117	+ 323.1 %	0.290	+ 35.6 %	0.090

Horizon = 192에서의 실험 결과

$$\Delta \text{RMSE \%} = \frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}} \times 100 \%,$$

양수 (+) 면 성능 악화 (오차 커짐), 음수 (-) 면 성능 개선 (오차 감소) 를 의미함

|ΔACF day|, |ΔACF week| 는 Train 구간에서 ACF 감소의 평균 절댓값 크기로, 값이 클수록 해당 주기 구조를 더 많이 망가뜨렸다는 의미

3. Experiment

Data-Level Analysis: H1

Datasets	Baseline RMSE (mean ± std)	Shuffle day Δ RMSE %	Shuffle day ΔACF day	Shuffle week Δ RMSE %	Shuffle week ΔACF week	Sdiff day Δ RMSE %	Sdiff day ΔACF day	Sdiff week Δ RMSE %	Sdiff week ΔACF week
ETTh1	3.311 ± 0.014	+ 40.2 %	0.154	+ 39.6 %	0.142	+ 78.0 %	0.456	+ 28.5 %	0.296
ETTh2	3.728 ± 0.018	+ 7.0 %	0.027	+ 3.8 %	0.020	+ 169.2 %	0.623	+ 33.4 %	0.355
ETTm1	3.230 ± 0.008	+ 44.1 %	0.152	+ 42.8 %	0.141	+ 49.0 %	0.448	+ 24.4 %	0.298
ETTm2	3.675 ± 0.018	+ 7.5 %	0.028	+ 4.4 %	0.021	+ 162.1 %	0.614	+ 32.2 %	0.341
Weather	42.097 ± 0.318	+ 29.8 %	0.125	+ 27.2 %	0.117	+ 307.7 %	0.292	+ 31.8 %	0.090

Horizon = 336에서의 실험 결과

Datasets	Baseline RMSE (mean ± std)	Shuffle day Δ RMSE %	Shuffle day ΔACF day	Shuffle week Δ RMSE %	Shuffle week ΔACF week	Sdiff day Δ RMSE %	Sdiff day ΔACF day	Sdiff week Δ RMSE %	Sdiff week ΔACF week
ETTh1	3.168 ± 0.029	+ 40.8 %	0.153	+ 40.3 %	0.139	+ 69.4 %	0.450	+ 20.7 %	0.291
ETTh2	3.649 ± 0.010	+ 7.0 %	0.027	+ 5.0 %	0.020	+ 213.1 %	0.629	+ 34.3 %	0.372
ETTm1	3.355 ± 0.014	+ 43.3 %	0.153	+ 42.5 %	0.140	+ 47.5 %	0.448	+ 26.0 %	0.300
ETTm2	4.100 ± 0.016	+ 5.9 %	0.028	+ 2.0 %	0.020	+ 149.9 %	0.614	+ 38.1 %	0.337
Weather	41.685 ± 0.169	+ 27.0 %	0.124	+ 23.9 %	0.115	+ 310.7 %	0.290	+ 31.3 %	0.086

Horizon = 720에서의 실험 결과

$$\Delta \text{RMSE \%} = \frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}} \times 100 \%,$$

양수 (+) 면 성능 악화 (오차 커짐), 음수 (-) 면 성능 개선 (오차 감소) 를 의미함

|ΔACF day|, |ΔACF week| 는 Train 구간에서 ACF 감소의 평균 절댓값 크기로, 값이 클수록 해당 주기 구조를 더 많이 망가뜨렸다는 의미

3. Experiment

Data-Level Analysis: H2

- H2 실험 결과, ETTm2, ETTh2 는 채널 간 상관관계보다 “같은 시점에서 함께 등장” 하는 시간 정합에서 제안 모델로 인한 이득이 발생하는 것으로 확인됨
- α Cross Stitch 로 인한 이득은 미약한 것으로 확인되어 미세 조정 수준에 그치는 것으로 보임
- 이때, $\Delta \text{RMSE} = \text{RMSE}(\text{Test}) - \text{RMSE}(\text{Baseline})$ 에 대한 결과값(+: 악화, -: 이득)

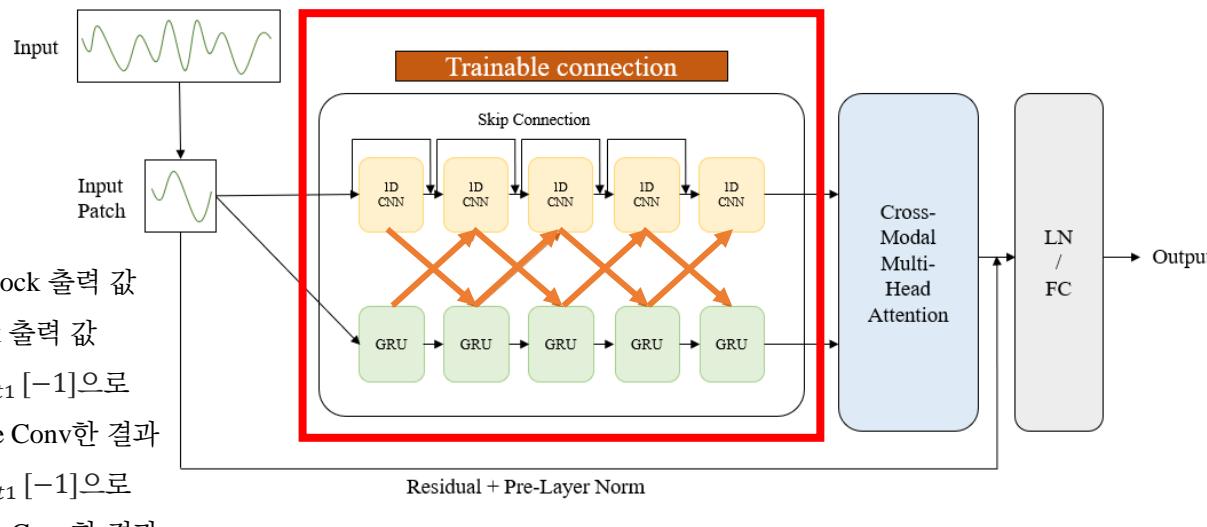
Datasets	Mean $ \rho $	ci-entropy	roll-week ΔRMSE	drop-corr-25% ΔRMSE	TAR	α gain
ETTm2	0.3245	0.7537	+ 0.576	+ 0.111	+ 0.465	+ 0.002
ETTh2	0.3246	0.7533	+ 0.228	-0.052	+ 0.280	+ 0.000
ETTm1	0.2243	0.7787	+ 0.232	+ 0.364	-0.132	-0.000
ETTh1	0.2221	0.7799	+ 0.081	+ 0.347	-0.266	-0.001

- Mean $|\rho|$: “채널들이 동시에 얼마나 같이 움직이는가?” 를 0 ~ 1 사이로 평균 낸 값
- ci-entropy (correlation-informed entropy): “변동이 소수의 공통 성분에 얼마나 모여 있는가?” 낮을 수록 한두 개 공통 패턴에 변동이 집중됨
- roll-week RMSE: 각 채널을 서로 다른 “최대 1주” 시간만큼 엇갈리게 만들어 동시성을 깨뜨린 뒤의 오차
- drop-corr-25% RMSE: 다른 채널들과 상관이 높은 상위 25% 채널을 삭제하고 학습했을 때의 RMSE
- TAR (Time-Alignment Reliance): 동시성 의존도 지표, $(\text{roll-week } \Delta \text{RMSE}) - (\text{drop-corr-25\% } \Delta \text{RMSE})$
- α gain: α 공유로 인한 순이득, $(\alpha \text{ Cross Stitch OFF}) - (\alpha \text{ Cross Stitch ON})$

3. Experiment

Data-Level Analysis: H3

- 제안 모델의 연산을 수식으로 표현하면 다음과 같이 진행됨
 - ✓ Input Patch $\rightarrow C_{out1} = 1dConv(\text{Input Patch}) \rightarrow C_h = \text{Dynamic Depth-Wise Conv}(C_{out1}, G_{out1}[-1])$
 - ✓ Input Patch $\rightarrow G_{out1} = \text{GRU}(\text{Input Patch}) \rightarrow G_h = \text{Dynamic Depth-Wise Conv}(G_{out1}, C_{out1}[-1])$
 - ❖ 이때 $C_{out1}[-1]$, $G_{out1}[-1]$ 은 각 Block의 마지막 시점 (조건 벡터)으로, 배치·채널·시점별로 Kernel을 변화하기 위해 필요한 정보
- H3 실험은 $C_{out1}[-1]$, $G_{out1}[-1]$ 을 조작함으로써 Dynamic Depth-Wise Conv 가 ETTm2, h2 의 성능 이득에 기여하였는지 판단하는 실험임



3. Experiment

Data-Level Analysis: H3

- H3 실험에서는 가설 검증을 위하여 다음의 4개 조작을 수행함
 - Freeze-First: 첫 배치에서 추출한 조건 벡터를 이후 전체에 고정
 - EMA: 조건 벡터 생성 시 새 정보에 대한 반영 비율을 줄여 변동성을 낮춤
 - Zero: 조건 벡터를 영벡터로 교체하여 조건성을 제거
 - Static Depth-Wise: 정적 채널별 고정 Kernel로 대체하여 조건 벡터를 사용하지 않음
- 실험 결과, 조건성을 강하게 약화(EMA/Zero)하거나 완전히 제거(Static DW) 하면 **ETTm2, h2**에서 성능이 크게 감소하였고, 긴 주기와 Horizon이 길수록 악화 폭이 더 커지는 것을 확인할 수 있었음
- **ETTm2, h2**는 일주기 정합이 Datasets 중 가장 강한 것으로 확인됨, 이러한 Data는 작은 위상·진폭 차이로 예측 오차가 달라질 수 있어 즉시 위상·진폭차를 보정하는 Dynamic Depth-Wise가 큰 이득을 만듦으로 보임

Datasets	Phase-Drift (↓= 정합↑)	SPAI (↑= 정합↑)	Δ RMSE % @ H = 720 (EMA)	Δ RMSE % @ H = 720 (Static-DW)
ETTh2	0.0417	0.8193	+ 18.5 %	+ 17.5 %
ETTm2	0.0208	0.8221	+ 8.8 %	+ 8.2 %
ETTh1	0.0833	0.6530	–	+ 3.3 %
ETTm1	0.0885	0.6496	–	+ 1.8 %
Weather	0.0486	0.7675	+ 4.3 %	+ 1.4 %

3. Experiment

Data-Level Analysis: H3: Phase-Drift

- **Phase-Drift** (작을 수록 정합 ↑): 오늘 파형 x_d 가 어제 파형 x_{d-1} 에 비해 얼마나 시계열 상 밀려 있는지 (Shift 되어 있는지) 를 정수 샘플 단위로 측정하는 도구
- 계산 순서는 다음과 같음
 - ① 하루 파형을 표준화하여 스케일 영향을 제거

$$\tilde{x}_d = \frac{x_d - \mu(x_d)}{\sigma(x_d) + \varepsilon}$$

- ② 모든 정수 Shift $l \in \{0, 1, \dots, T - 1\}$ 에 대해, 전일 파형을 원형 Shift roll(\tilde{x}_{d-1}, l) 하고 피어슨 상관을 계산

$$\rho(l) = \text{corr}(\tilde{x}_d, \text{roll}(\tilde{x}_{d-1}, l))$$

- ③ 최적 시프트 $\tilde{l} = \text{argmax}_l \rho(l)$ 로 상관이 최대가 되는 l 을 선택, 그 절대 크기를 하루 길이로 정규화함

$$\text{Phase-Drift}_d = \frac{|\tilde{l}|}{T} \in [0, 0.5]$$

- 예를 들어, ETTh (1시간 간격, T=24) = Phase-Drift = $\frac{1}{24} = 0.0417$; 하루 간 어긋남이 1시간이라는 의미
- 원형 Shift roll: 벡터를 원형으로 회전시켜 하루의 시작/끝을 넘어가도 자연스럽게 맞추도록 만듦

3. Experiment

Data-Level Analysis: H3: SPAI

- **SPAI** (Shape / Phase Alignment Index, 클 수록 정합↑): Shift 를 보정 후 오늘 파형 x_d 과 어제 파형 x_{d-1} 의 유사도 (상관) 을 측정하는 도구
- 계산 순서는 다음과 같음
 - ① 하루 파형을 표준화하여 스케일 영향을 제거

$$\tilde{x}_d = \frac{x_d - \mu(x_d)}{\sigma(x_d) + \varepsilon}$$

- ② 모든 정수 Shift $l \in \{0, 1, \dots, T - 1\}$ 에 대해, 전일 파형을 원형 Shift roll(\tilde{x}_{d-1}, l) 하고 피어슨 상관을 계산

$$\rho(l) = \text{corr}(\tilde{x}_d, \text{roll}(\tilde{x}_{d-1}, l))$$

- ③ 최적 시프트 $\tilde{l} = \text{argmax}_l \rho(l)$ 로 상관이 최대가 되는 l 을 선택하여 다음과 같은 계산을 취함

$$\text{SPAI}_d = \text{corr}(\tilde{x}_d, \text{roll}(\tilde{x}_{d-1}, \tilde{l})) \in [-1, 1]$$

- 원형 Shift roll: 벡터를 원형으로 회전시켜 하루의 시작/끝을 넘어가도 자연스럽게 맞추도록 만듦

3. Experiment

Data-Level Analysis: H3

Horizon	Baseline RMSE (mean \pm std)	Freeze-First Δ RMSE %	EMA Δ RMSE %	Zero Δ RMSE %	Static-DW Δ RMSE %
96	3.601 ± 0.011	+ 1.7 %	+ 8.7 %	+ 9.3 %	+ 8.7 %
192	3.653 ± 0.011	+ 1.5 %	+ 10.2 %	+ 10.0 %	+ 9.1 %
336	3.728 ± 0.020	+ 0.9 %	+ 8.6 %	+ 8.8 %	+ 7.7 %
720	3.644 ± 0.010	- 0.1 %	+ 18.5 %	+ 12.7 %	+ 17.5 %

ETTh2에서의 실험 결과

Horizon	Baseline RMSE (mean \pm std)	Freeze-First Δ RMSE %	EMA Δ RMSE %	Zero Δ RMSE %	Static-DW Δ RMSE %
96	2.959 ± 0.005	+ 0.3 %	+ 6.3 %	+ 5.9 %	+ 3.9 %
192	3.338 ± 0.019	- 0.1 %	+ 7.2 %	+ 6.8 %	+ 5.1 %
336	3.683 ± 0.017	- 0.6 %	+ 8.3 %	+ 7.5 %	+ 6.0 %
720	4.098 ± 0.014	- 0.5 %	+ 8.8 %	+ 8.0 %	+ 8.2 %

ETTm2에서의 실험 결과

Horizon	Baseline RMSE (mean \pm std)	Freeze-First Δ RMSE %	EMA Δ RMSE %	Zero Δ RMSE %	Static-DW Δ RMSE %
96	39.718 ± 0.135	+ 0.4 %	+ 6.7 %	+ 6.2 %	+ 2.9 %
192	40.850 ± 0.193	+ 0.3 %	+ 7.1 %	+ 6.1 %	+ 3.5 %
336	42.127 ± 0.270	- 0.1 %	+ 5.1 %	+ 4.4 %	+ 2.6 %
720	41.678 ± 0.179	+ 0.3 %	+ 4.3 %	+ 5.2 %	+ 1.4 %

$$\Delta \text{RMSE \%} = \frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}} \times 100 \%, \quad \text{Weather에서의 실험 결과}$$

양수 (+) 면 성능 악화(오차 커짐), 음수 (-) 면 성능 개선(오차 감소)를 의미함

3. Experiment

Data-Level Analysis: H3

Horizon	Baseline RMSE (mean \pm std)	Freeze-First Δ RMSE %	EMA Δ RMSE %	Zero Δ RMSE %	Static-DW Δ RMSE %
96	3.055 ± 0.008	+ 0.5 %	—	—	+ 1.5 %
192	3.185 ± 0.015	+ 0.6 %	—	—	+ 0.7 %
336	3.311 ± 0.014	+ 0.5 %	—	—	+ 0.6 %
720	3.168 ± 0.029	+ 0.9 %	—	—	+ 3.3 %

ETTh1에서의 실험 결과

Horizon	Baseline RMSE (mean \pm std)	Freeze-First Δ RMSE %	EMA Δ RMSE %	Zero Δ RMSE %	Static-DW Δ RMSE %
96	2.818 ± 0.007	- 1.3 %	—	—	+ 2.5 %
192	3.079 ± 0.006	- 0.6 %	—	—	+ 1.8 %
336	3.230 ± 0.008	- 0.3 %	—	—	+ 2.0 %
720	3.335 ± 0.014	- 0.1 %	—	—	+ 1.8 %

ETTm1에서의 실험 결과

$$\Delta \text{RMSE \%} = \frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}} \times 100 \%,$$

양수 (+) 면 성능 악화 (오차 커짐), 음수 (-) 면 성능 개선 (오차 감소) 를 의미함

3. Experiment

Data-Level Analysis: H4

- H4 실험 결과, 모든 ETT 계열 Data에서 SNR 을 낮출수록 RMSE 가 증가하고 그 관계가 높은 선형성을 보임
- SNR 을 낮추었을 때, ETTm2 가 ETTm1 보다 하락이 크게 나타났으나 ETTh2 는 ETTh1 보다 하락폭이 작았음
- $\text{Elasticity} = \frac{\Delta \text{RMSE \%}}{\Delta \text{SNR \%}}$ (값이 클수록 SNR 약화에 더 민감) 가 ETTm2, ETTh2 에서 양 (+) 이며 선형 적합도 (R^2) 또한 유의미하여 제안 모델이 Data 의 저주파 골격 (추세 · 주기) 에 실질적으로 의존한다고 볼 수 있음
- ETTm1, ETTh1 은 Baseline SNR 이 상대적으로 낮고 하루 주기 위상 안정성 (ACF@day) 이 낮아 제안 모델로 인한 이득이 떨어지고, Weather 는 채널 간 이질성이 강하여 제안 모델로 인한 이득이 크게 드러나지 않는 것으로 보임 (아래 표의 Weather SNR 은 채널 중앙값을 사용함)

Datasets	P_{day}	Vars	Baseline SNR_{time}	ACF@day
ETTm2	96	7	22.7629	0.8243
ETTh2	24	7	22.8323	0.8231
ETTm1	96	7	6.4908	0.7991
ETTh1	24	7	6.5134	0.8021
Weather	144	21	7.9551	0.5904

$$\Delta \text{RMSE \%} = \frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}} \times 100 \%, \text{ 양수 (+) 면 성능 악화 (오차 커짐), 음수 (-) 면 성능 개선 (오차 감소) 를 의미함}$$

ACF@day: 하루 주기(P_{day} /ETTm: 96 Steps, ETTh: 24 Steps, Weather: 144 Steps) 시차 자기 상관으로, 하루 주기의 위상 안정성을 반영함

3. Experiment

Data-Level Analysis: H4

- Weather Dataset 의 SNR 상위 6개, 하위 6개 채널 (변수)에 대한 정보는 다음과 같음
 - 저주파 지배군은 SNR 이 매우 크고 ACF@day 도 높아 명확한 일 주기 문맥을 형성하나, 고주파/불규칙 지배군은 SNR 이 매우 작고 ACF@day 도 낮아 주기 위상 안정성이 약함
 - 곧, 채널 간 이질성이 강하다고 할 수 있음 (표의 SNR 은 단순 산술평균수치)
- 이때 저주파 분산 비율 (LF ratio) 는 $\frac{SNR}{1+SNR}$ 과 같음

Variable	SNR_{time}	LF ratio	ACF@day	Variable	SNR_{time}	LF ratio	ACF@day
p (mbar)	3507.32	0.9997	0.9849	wv (m/s)	0.0446	0.0427	0.0005
rho (g/m^3)	376.11	0.9973	0.9176	rain (mm)	0.4361	0.3037	0.0820
Tlog (degC)	321.42	0.9969	0.8764	wd (deg)	1.1339	0.5314	0.3597
Tpot (K)	300.76	0.9967	0.9002	raining (s)	1.6069	0.6164	0.2983
T (degC)	295.45	0.9966	0.8987	max. PAR	2.4645	0.7114	0.3297
Tdew (degC)	285.34	0.9965	0.9659	OT	4.1571	0.8061	0.2019

(좌) 저주파 지배군, (우) 고주파 / 불규칙 지배군

$$\Delta \text{RMSE \%} = \frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}} \times 100 \%, \text{ 양수 (+) 면 성능 악화 (오차 커짐), 음수 (-) 면 성능 개선 (오차 감소) 를 의미함}$$

ACF@day: 하루 주기(P_{day} /ETTm: 96 Steps, ETTh: 24 Steps, Weather: 144 Steps) 시차 자기 상관으로, 하루 주기의 위상 안정성을 반영함

3. Experiment

Data-Level Analysis: H4

Datasets	Horizon	Baseline RMSE (mean ± std)	SNR-down RMSE (mean ± std)	Δ RMSE %	Elasticity	Fit R^2
ETTm2	96	2.959 ± 0.005	3.125 ± 0.099	+ 5.6 %	0.124	0.851
	192	3.338 ± 0.019	3.478 ± 0.089	+ 4.2 %	0.094	0.791
	336	3.683 ± 0.017	3.786 ± 0.069	+ 2.8 %	0.062	0.710
	720	4.098 ± 0.014	4.159 ± 0.058	+ 1.5 %	0.041	0.590
Datasets	Horizon	Baseline RMSE (mean ± std)	SNR-down RMSE (mean ± std)	Δ RMSE %	Elasticity	Fit R^2
ETTh2	96	3.601 ± 0.011	3.664 ± 0.061	+ 1.7 %	0.050	0.646
	192	3.653 ± 0.011	3.719 ± 0.062	+ 1.8 %	0.050	0.621
	336	3.728 ± 0.020	3.806 ± 0.072	+ 2.1 %	0.056	0.621
	720	3.644 ± 0.010	3.730 ± 0.083	+ 2.4 %	0.066	0.619
Datasets	Horizon	Baseline RMSE (mean ± std)	SNR-down RMSE (mean ± std)	Δ RMSE %	Elasticity	Fit R^2
Weather	96	39.718 ± 0.135	40.208 ± 0.307	+ 1.2 %	0.023	0.605
	192	40.850 ± 0.193	41.107 ± 0.249	+ 0.6 %	0.014	0.424
	336	42.127 ± 0.270	42.234 ± 0.261	+ 0.3 %	0.003	0.015
	720	41.678 ± 0.179	42.139 ± 0.344	+ 1.1 %	0.023	0.546

$\Delta \text{RMSE \%} = \frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}} \times 100 \%$, 양수 (+) 면 성능 악화 (오차 커짐), 음수 (-) 면 성능 개선 (오차 감소) 를 의미함

Elasticity = $\frac{\Delta \text{RMSE \%}}{\Delta \text{SNR \%}}$, SNR $\circ|$ 1% 감소할 때 RMSE 가 몇 % 증가하는가를 나타내는 지표

SNR-down RMSE 는 모든 변조에서의 총 평균값을 의미함

3. Experiment

Data-Level Analysis: H4

Datasets	Horizon	Baseline RMSE (mean \pm std)	SNR-down RMSE (mean \pm std)	Δ RMSE %	Elasticity	Fit R^2
ETTm1	96	2.817 ± 0.007	2.898 ± 0.049	+ 2.9 %	0.062	0.802
	192	3.081 ± 0.006	3.157 ± 0.048	+ 2.5 %	0.054	0.781
	336	3.230 ± 0.008	3.315 ± 0.050	+ 2.6 %	0.056	0.814
	720	3.335 ± 0.014	3.408 ± 0.052	+ 2.2 %	0.052	0.715
Datasets	Horizon	Baseline RMSE (mean \pm std)	SNR-down RMSE (mean \pm std)	Δ RMSE %	Elasticity	Fit R^2
ETTh1	96	3.056 ± 0.008	3.210 ± 0.099	+ 5.1 %	0.114	0.808
	192	3.186 ± 0.015	3.346 ± 0.103	+ 5.0 %	0.113	0.798
	336	3.310 ± 0.015	3.471 ± 0.107	+ 4.9 %	0.111	0.780
	720	3.164 ± 0.032	3.374 ± 0.151	+ 6.6 %	0.157	0.740

Δ RMSE % = $\frac{(\text{Test RMSE} - \text{Baseline RMSE})}{\text{Baseline RMSE}}$ $\times 100\%$, 양수 (+) 면 성능 악화 (오차 커짐), 음수 (-) 면 성능 개선 (오차 감소) 를 의미함

Elasticity = $\frac{\Delta \text{RMSE \%}}{\Delta \text{SNR \%}}$, SNR $\circ|$ 1% 감소할 때 RMSE 가 몇 % 증가하는가를 나타내는 지표

SNR-down RMSE 는 모든 변조에서의 총 평균값을 의미함

3. Experiment

Data-Level Analysis: H5

- H5 실험 결과에 따르면 Co-Attention Block 을 Concat 으로 교체하면 특히 ETTm2, ETTh2 의 모든 Horizon에서 RMSE 가 크게 하락하고 CC (Correlation Coefficient) 가 소폭 하락함
- 데이터 분석에 따르면 ETTm2, ETTh2 는 “어제 / 지난주 같은 시간”의 정보가 유용하고 ($ACF_P \uparrow$) 데이터 흐름이 비교적 안정적으로 재현되며 ($Phase_var_P \downarrow$) 여러 변수가 그 Lag 로 함께 움직임 ($Cross_sync_P \uparrow$)
- Co-Attention 은 이때 1D Conv 토큰 (Query) 과 GRU 문맥 (Key / Value) 을 선택적으로 가중해 결합하여 CC 저하를 억제하고 RMSE 악화를 줄이거나, Weather, ETTm1, ETTh1 은 주 Lag 정합 ($WAI < 0$ 혹은 전체 지수가 낮아 선택적 결합을 해도 이득이 작거나 혼재함

Datasets	ACF _P		Align_index _P		Cross_sync _P		Phase_var _P	
	DAI	WAI	DAI	WAI	DAI	WAI	DAI	WAI
ETTm2	0.8892	0.7349	0.6389	0.0937	0.2009	0.1673	0.4512	0.8085
ETTh2	0.8798	0.7271	0.6236	0.0625	0.2005	0.1673	0.4567	0.8319
ETTm1	0.7903	0.6211	0.3743	-0.0937	0.1770	0.1530	0.5930	0.8677
ETTh1	0.7331	0.5699	0.3149	-0.1581	0.1753	0.1514	0.5936	0.8794
Weather	0.6091	0.4683	0.2759	-0.2967	0.1104	0.0987	0.4436	0.8638

각 Datasets 에 대한 데이터 분석 결과

3. Experiment

Data-Level Analysis: H5: ACF_P

- ACF_p : “Lag P (하루 / 일주일 전) 의 과거와 현재가 얼마나 닮았는가” 를 수치화하는 지표로, 값이 클수록 해당 주기 Lag 에서 패턴이 반복된다는 의미
- ACF_p 가 클 수록 Lag P 가 유용 (과거 P 시점의 정보가 현재와 잘 맞닿음)
- 일 / 주 ACF_p 가 높다는 것은 “어제 / 지난주 같은 시간” 의 패턴이 잘 반복된다는 의미
- 본 실험에서는 채널 별 $ACF_k^{(c)}$ 를 구하여 밴드 평균값을 낸 뒤, 다시 채널 평균값을 구함

$$ACF_k^{(c)} = \frac{\sum_{t=k}^{T-1} (x_t^{(c)} - \bar{x}^{(c)})(x_{t-k}^{(c)} - \bar{x}^{(c)})}{\sum_{t=0}^{T-1} (x_t^{(c)} - \bar{x}^{(c)})^2 + \varepsilon}, \quad \bar{x}^{(c)} = \frac{1}{T} \sum_{t=0}^{T-1} x_t^{(c)}$$

$$\overline{ACF}_P^{(c)} = \frac{1}{2b+1} \sum_{k=P-b}^{P+b} ACF_k^{(c)}$$

$$ACF_P = \frac{1}{C} \sum_{c=1}^C \overline{ACF}_P^{(c)}$$

T: 시계열의 총 길이 / t: 시간 인덱스 / c: 채널 인덱스

P: Lag, 얼마나 과거를 참고하느냐라는 시간 간격

$x_t^{(c)}$: 채널 c의 시간 t에서의 관측 값 / $\bar{x}^{(c)}$: 채널 c에서의 시간에 따른 관측 값 평균

3. Experiment

Data-Level Analysis: H5: Phase_var_P

- Phase_var_P: 일 / 주를 길이 P 인 원으로 전환하여 피크가 발생한 시점을 각도로 전환하여 측정
 - Phase_var_P 가 작다: 피크 타이밍이 안정 → 일 / 주 비슷한 시각에 피크가 발생함 (위상 안정, 변동 작음)

① 채널 c 의 시계열 $x^{(c)}$ 를 길이 P 로 연속 블록 ($B_i^{(c)}$) 으로 분할

$$B_i^{(c)} = \{x_{iP}^{(c)}, x_{iP+1}^{(c)}, \dots, x_{(i+1)P-1}^{(c)}\}, \quad i = 0, \dots, \frac{T}{P} - 1$$

- 예를 들어, Data: ETTh 10일 ($\rightarrow T: 240$), P: 24, c: “전력부하” 일 때 $i = 2$ (세 번째 날) 이면
 - $B_2^{(c)} = \{x_{48}^{(c)}, x_{49}^{(c)}, \dots, x_{71}^{(c)}\}$ 로 P 개의 시점을 포함하는 블록이 생성됨
 - ($x_{48}^{(c)}$ 은 ‘세 번째 날의 0시의 채널 “전력부하“ 의 정보라고 이해할 수 있음)
- ② 각 블록에서 최댓값 시점 τ 을 추출 (결측은 건너뜀)

$$p_i^{(c)} = \operatorname{argmax}_{0 \leq \tau < P} x_{iP+\tau}^{(c)} \in \{0, \dots, P-1\}$$

T : 시계열 길이 / B : 밴드 폭 / c : 채널(변수) 인덱스

P : Lag 길이(얼마나 과거를 참고하느냐 라는 시간 간격)

3. Experiment

Data-Level Analysis: H5: Phase_var_P

- Phase_var_P: 일 / 주를 길이 P 인 원으로 전환하여 피크가 발생한 시점을 각도로 전환하여 측정
 - Phase_var_P 가 작다: 피크 타이밍이 안정 → 일 / 주 비슷한 시각에 피크가 발생함 (위상 안정, 변동 작음)
- ③ 시점 $p_i^{(c)}$ 을 각도로 변환함

$$\theta_i^{(c)} = \frac{2\pi}{P} p_i^{(c)} \text{ (radian)}$$

- ④ 블록 별 각도들의 평균 방향과 집중 정도를 구함

$$C^{(c)} = \frac{1}{N} \sum_i \cos \theta_i^{(c)}, \quad S^{(c)} = \frac{1}{N} \sum_i \sin \theta_i^{(c)}, \quad R^{(c)} = \sqrt{(C^{(c)})^2 + (S^{(c)})^2}$$

- ⑤ 원형 분산으로 바꾸고 채널 평균을 계산함

$$\text{Phase_var}_P = \frac{1}{C} \sum_{c=1}^C 1 - R^{(c)}$$

3. Experiment

Data-Level Analysis: H5: Cross_sync_P

- Cross_sync_P: 여러 변수쌍이 같은 Lag P에서 동시에 움직이는지를 확인하는 지표 (동시에 움직이면 그 Lag를 강조할 전역적 근거가 큼)

① 각 채널 별 평균을 제거하여 스케일 / 오프셋의 영향을 줄임

$$\tilde{x}_t^{(c)} = x_t^{(c)} - \bar{x}^{(c)}, \quad \bar{x}^{(c)} = \frac{1}{T} \sum_{t=0}^{T-1} x_t^{(c)}$$

② 정규화 상수를 계산

$$\|\tilde{x}^{(c)}\| = \sqrt{\sum_{t=0}^{T-1} (\tilde{x}_t^{(c)})^2 + \varepsilon}$$

③ 모든 채널 전 쌍 (i, j), i < j 에 대해 쌍 별 교차상관 (Lag P) 을 구함

$$\text{xcorr}_P^{(i,j)} = \frac{\sum_{t=P}^{T-1} \tilde{x}_t^{(i)} \tilde{x}_{t-P}^{(j)}}{\|\tilde{x}^{(i)}\| \|\tilde{x}^{(j)}\|}$$

④ 모든 쌍을 사용해 평균을 구함

$$\text{cross_sync}_P = \frac{1}{\binom{C}{2}} \sum_{i < j} \text{xcorr}_P^{(i,j)}$$

3. Experiment

Data-Level Analysis: H5: Align_index_P, DAI, WAI

- Align_index_P: 하나의 Lag P에서 (i) 유용성, (ii) 타이밍 안정성, (iii) 다변수 동조화를 하나의 수치로 요약한 지표
- 크기가 양수일수록 해당 Lag P가 유용, 안정, 동조하여 Lag를 골라 쓰는 결합이 맞아 떨어질 환경임
- 크기가 음수일수록 해당 Lag P는 덜 유용 / 불안정하여 그 Lag를 강조해도 효과가 작거나 혼재할 소지 있음

$$\text{Align_index}_P = \frac{\text{ACF}_P}{\text{유용성}} - \frac{\text{Phase_var}_P}{\text{타이밍 안정}} + \frac{\text{Cross_sync}_P}{\text{동조화}}$$

- **DAI** (Day Alignment Index): 하루 Lag에서의 Align_index

$$\text{DAI} = \text{ACF}_{P_{day}} - \text{Phase_var}_{P_{day}} + \text{Cross_sync}_{P_{day}}$$

- **WAI** (Week Alignment Index): 일주 Lag에서의 Align_index

$$\text{WAI} = \text{ACF}_{P_{week}} - \text{Phase_var}_{P_{week}} + \text{Cross_sync}_{P_{week}}$$

3. Experiment

Data-Level Analysis: H5

Datasets	Horizon	Model	Avg. RMSE (mean \pm std)	Δ RMSE vs Baseline	Avg. CC (mean \pm std)	Δ CC vs Baseline
ETTm2	96	Baseline (Co-Attn)	2.960 ± 0.007	–	0.9856 ± 0.0001	–
		Ablation (Concat)	3.181 ± 0.056	+ 7.5 %	0.9834 ± 0.0006	-0.0022
	192	Baseline (Co-Attn)	3.345 ± 0.008	–	0.9817 ± 0.0001	–
		Ablation (Concat)	3.633 ± 0.059	+ 8.6 %	0.9784 ± 0.0007	-0.0032
	336	Baseline (Co-Attn)	3.693 ± 0.020	–	0.9777 ± 0.0003	–
		Ablation (Concat)	4.048 ± 0.056	+ 9.6 %	0.9732 ± 0.0008	-0.0045
	720	Baseline (Co-Attn)	4.097 ± 0.011	–	0.9727 ± 0.0001	–
		Ablation (Concat)	4.361 ± 0.013	+ 6.4 %	0.9691 ± 0.0002	-0.0036

Datasets	Horizon	Model	Avg. RMSE (mean \pm std)	Δ RMSE vs Baseline	Avg. CC (mean \pm std)	Δ CC vs Baseline
ETTh2	96	Baseline (Co-Attn)	3.603 ± 0.011	–	0.9807 ± 0.0001	–
		Ablation (Concat)	3.823 ± 0.053	+ 6.1 %	0.9779 ± 0.0005	-0.0029
	192	Baseline (Co-Attn)	3.653 ± 0.011	–	0.9812 ± 0.0001	–
		Ablation (Concat)	3.937 ± 0.014	+ 7.8 %	0.9776 ± 0.0001	-0.0037
	336	Baseline (Co-Attn)	3.728 ± 0.019	–	0.9810 ± 0.0001	–
		Ablation (Concat)	4.035 ± 0.034	+ 8.2 %	0.9771 ± 0.0003	-0.0039
	720	Baseline (Co-Attn)	3.643 ± 0.015	–	0.9809 ± 0.0001	–
		Ablation (Concat)	3.893 ± 0.013	+ 9.3 %	0.9774 ± 0.0001	-0.0036

3. Experiment

Data-Level Analysis: H5

Datasets	Horizon	Model	Avg. RMSE (mean \pm std)	Δ RMSE vs Baseline	Avg. CC (mean \pm std)	Δ CC vs Baseline
ETTm1	96	Baseline (Co-Attn)	2.817 ± 0.007	–	0.8599 ± 0.0008	–
		Ablation (Concat)	2.829 ± 0.012	+ 0.4 %	0.8584 ± 0.0012	-0.0015
	192	Baseline (Co-Attn)	3.081 ± 0.006	–	0.8315 ± 0.0008	–
		Ablation (Concat)	3.093 ± 0.014	+ 0.4 %	0.8299 ± 0.0016	-0.0015
	336	Baseline (Co-Attn)	3.230 ± 0.008	–	0.8142 ± 0.0011	–
		Ablation (Concat)	3.265 ± 0.021	+ 1.1 %	0.8105 ± 0.0023	-0.0038
	720	Baseline (Co-Attn)	3.335 ± 0.014	–	0.8058 ± 0.0018	–
		Ablation (Concat)	3.351 ± 0.027	+ 0.5 %	0.8044 ± 0.0030	-0.0014
Datasets	Horizon	Model	Avg. RMSE (mean \pm std)	Δ RMSE vs Baseline	Avg. CC (mean \pm std)	Δ CC vs Baseline
ETTh1	96	Baseline (Co-Attn)	3.056 ± 0.008	–	0.8406 ± 0.0009	–
		Ablation (Concat)	3.134 ± 0.024	+ 2.6 %	0.8310 ± 0.0030	-0.0096
	192	Baseline (Co-Attn)	3.186 ± 0.015	–	0.8248 ± 0.0017	–
		Ablation (Concat)	3.293 ± 0.030	+ 3.4 %	0.8101 ± 0.0036	-0.0147
	336	Baseline (Co-Attn)	3.310 ± 0.015	–	0.8056 ± 0.0019	–
		Ablation (Concat)	3.430 ± 0.025	+ 3.6 %	0.7863 ± 0.0032	-0.0193
	720	Baseline (Co-Attn)	3.164 ± 0.032	–	0.8145 ± 0.0032	–
		Ablation (Concat)	3.330 ± 0.041	+ 5.3 %	0.7882 ± 0.0053	-0.0263

3. Experiment

Data-Level Analysis: H5

Datasets	Horizon	Model	Avg. RMSE (mean \pm std)	Δ RMSE vs Baseline	Avg. CC (mean \pm std)	Δ CC vs Baseline
Weather	96	Baseline (Co-Attn)	39.678 ± 0.136	–	0.9929 ± 0.0000	–
		Ablation (Concat)	40.289 ± 0.129	+ 1.5 %	0.9926 ± 0.0001	-0.0002
	192	Baseline (Co-Attn)	40.823 ± 0.186	–	0.9925 ± 0.0001	–
		Ablation (Concat)	41.099 ± 0.185	+ 0.7 %	0.9924 ± 0.0001	-0.0001
	336	Baseline (Co-Attn)	42.055 ± 0.301	–	0.9920 ± 0.0001	–
		Ablation (Concat)	41.831 ± 0.102	-0.5 %	0.9921 ± 0.0000	-0.0001
	720	Baseline (Co-Attn)	41.677 ± 0.176	–	0.9921 ± 0.0001	–
		Ablation (Concat)	41.805 ± 0.116	+ 0.3 %	0.9921 ± 0.0000	-0.0000

3. Experiment

Data-Level Analysis: H6

- H6 실험 결과, **RevIN** (Reversible Instance Normalization) 을 OFF 하면 대부분의 Data에서 성능이 크게 하락하는 것을 확인할 수 있었음
- ETT (ETTm1, m2, h1, h2) Datasets에서 RevIN OFF 시 RMSE가 약 33% 증가(악화)하였으며, Weather Dataset에서 약 22 ~ 25% 증가(악화)하는 것을 확인할 수 있었음
- **RevIN**이 분포 드리프트를 줄인다는 가설은 일부 지지되는 것으로 확인됨
 - 분포 드리프트 관점에서, μ (Mean) 정렬은 폭넓게 작용하여 성능 개선의 주된 원인으로 생각됨
 - 분포 드리프트 관점에서, σ^2 (Variance) 정렬은 Datasets에 의존하되 예측 오차는 줄여 실용 관점에서 이득은 유지하는 것으로 생각됨

Datasets	ΔRMSE (OFF- ON)	$\Delta \mu_s^{abs}$ (OFF- ON)	$\Delta \sigma_s^{2abs}$ (OFF- ON)
ETTh1	1.4481	-0.4622	-3.6988
ETTm1	0.3436	0.6808	0.7271
ETTh2	2.7815	1.1541	4.6218
ETTm2	2.6948	1.5792	20.6237
Weather	14.5343	5.6061	-977.9211

Horizon = 720에서의 분포 드리프트 결과
(값이 양수면 RevIN이 드리프트를 감소시켰다는 의미)

3. Experiment

Data-Level Analysis: H6

- Datasets에 대해, Validation / Test 구간 Data를 길이 H(Horizon)의 Window로 잘라 각 창의 평균(μ_w)과 분산(σ_w)이 Train 구간의 평균, 분산에 비해 얼마나 멀어졌는지를 측정하여 데이터의 드리프트 정도를 확인함
- ETTm2, ETTh2는 타 Data에 비해 표준편차 1배 이상 Train 구간 평균에 비해 이동했으며, 분산 또한 기준 대비 80 ~ 90 % 변동하는 고 드리프트 Data임이 확인됨
- 그러나, 개별 수준 상관과 드리프트 감소량을 보았을 때 ETTm2는 드리프트 감소가 클 수록 RMSE 개선도 큰 것으로 보이지만 ETTh2는 드리프트 감소와 RMSE 개선 간의 상관이 낮았음
 - μ 정렬 외에도 RevIN의 학습 안정화 / 스케일 일관화를 통해 이득을 낸 것으로 추정

Horizon	ΔRMSE (mean \pm std)	$\Delta \mu$ drift (abs) (mean \pm std)	$\Delta \sigma^2$ drift (abs) (mean \pm std)	Horizon	ΔRMSE (mean \pm std)	$\Delta \mu$ drift (abs) (mean \pm std)	$\Delta \sigma^2$ drift (abs) (mean \pm std)
96	2.45 ± 1.34	0.84 ± 0.60	18.24 ± 14.76	96	2.36 ± 1.05	0.46 ± 0.39	-3.21 ± 11.42
192	3.00 ± 0.75	1.28 ± 0.43	23.62 ± 8.11	192	2.27 ± 0.90	0.17 ± 0.37	-8.55 ± 6.43
336	2.97 ± 0.13	1.48 ± 0.35	25.69 ± 9.49	336	2.33 ± 0.77	0.39 ± 0.48	-14.27 ± 6.77
720	2.69 ± 0.10	1.58 ± 0.34	20.62 ± 10.44	720	2.78 ± 0.33	1.15 ± 0.64	4.62 ± 17.26

(좌) ETTm2, (우) ETTh2에 대한 드리프트 감소량 표
(RevIN OFF – RevIN ON 값, +는 ON이 드리프트를 줄였다, -는 드리프트를 늘렸다는 의미)

3. Experiment

Data-Level Analysis: H6

- 공통 표기
 - 다변수 시계열 $x_t \in \mathbb{R}^C$ (채널 / 변수 수 C)
 - 훈련 구간 (60% Data)에서 채널별 기준 통계를 계산한 값은 다음과 같음

$$\mu_{0,c} = \text{mean}_{t \in \text{train}}(x_{t,c}), \quad \sigma_{0,c} = \text{std}_{t \in \text{train}}(x_{t,c}), \quad v_{0,c} = \sigma_{0,c}^2$$

- Validation / Test 구간을 길이 H (Horizon)의 Window로 분할, Window index = 1, ..., W 라 하고, Window w에서의 채널별 평균 / 분산은 다음과 같음

$$\mu_{w,c} = \text{mean}_{t \in w}(x_{t,c}), \quad v_{w,c} = \text{var}_{t \in w}(x_{t,c})$$

- 각 Window에서 채널 평균으로 집계한 값을 $d_\mu(w), d_v(w)$ 라 설정함
- 수치 안정화를 위해 분모에는 $\varepsilon \approx 10^{-9}$ 를 더함

3. Experiment

Data-Level Analysis: H6: μ drift (평균, SD)

- μ drift (평균, SD) – 평균이 기준에서 몇 표준편차만큼 이동했는지 전반적 수준을 측정하는 지표
- 훈련 기준 평균 μ_0 에서 얼마나 떨어져 있는지를 표준편차 σ_0 단위로 측정함
 - H6 실험 표의 값은 여러 Window 의 값을 평균한 값으로, 전체 구간의 전반적 평균 이동 정도를 나타냄
 - μ drift (평균, SD) 값이 1.0 이면 Validation / Test Window 의 평균이 Train 기준 평균에서 표준편차 1배 만큼 떨어져 있음을 의미함

$$d_\mu(w) = \frac{1}{C} \sum_{c=1}^C \frac{|\mu_{w,c} - \mu_{0,c}|}{\sigma_{0,c} + \varepsilon}$$

$$\mu \text{ drift} = \frac{1}{W} \sum_{w=1}^W d_\mu(w)$$

3. Experiment

Data-Level Analysis: H6: μ drift (p95, SD)

- μ drift (p95, SD) – 평균 이동이 가장 심한 5% Window 의 이동 크기를 측정하는 지표
- 훈련 기준 평균 μ_0 에서 얼마나 떨어져 있는지를 표준편차 σ_0 단위로 측정함
 - H6 실험 표의 값은 여러 Window 의 값을 평균한 값으로, 전체 구간의 전반적 평균 이동 정도를 나타냄
 - 전체 Window 중 상위 5%에서 평균이 가장 많이 움직인 구간의 스케일 수치
 - μ drift (p95, SD) = 1.55 이면 Validation / Test 상위 5% Window 의 평균이 Train 기준 평균에서 표준편차 1.55배 만큼 떨어져 있음을 의미함

$$d_\mu(w) = \frac{1}{C} \sum_{c=1}^C \frac{|\mu_{w,c} - \mu_{0,c}|}{\sigma_{0,c} + \epsilon}$$

$$\mu \text{ drift (p95)} = \text{quantile}_{0.95}(\{d_\mu(w)\}_{w=1}^W)$$

3. Experiment

Data-Level Analysis: H6: σ^2 drift (평균, 상대)

- σ^2 drift (평균, 상대) – Train 기준 분산 v_0 대비 Window 의 분산 v_w 가 얼마나 상대적으로 달라졌는지를 측정하는 무단위 비율 지표
 - σ^2 drift (평균, 상대) = 0.80 이면 분산이 Train 기준 대비 80% 만큼 달라졌다는 의미임
 - 이 값이 클수록 변동성 스케일 자체가 크게 흔들리는 Data 라는 의미임

$$d_v(w) = \frac{1}{C} \sum_{c=1}^C \frac{|v_{w,c} - v_{0,c}|}{v_{0,c} + \epsilon} \quad (v_{0,c} = \sigma_{0,c}^2)$$

$$\sigma^2 \text{ drift} = \frac{1}{W} \sum_{w=1}^W d_v(w)$$

3. Experiment

Data-Level Analysis: H6: σ^2 drift (p95, 상대)

- σ^2 drift (p95, 상대) – 분산 변화가 가장 심한 5% Window 의 크기를 측정하는 무단위 비율 지표
 - 전체 Window 중 상위 5%에서 분산 스케일이 얼마나 크게 변했는지를 확인하는 수치
 - σ^2 drift (p95, 상대) = 1.20 이면 분산이 Train 기준 대비 120% 만큼 달라졌다는 의미임
 - 급격한 Regime 전환 / 이상 변동을 포착하는 지표

$$d_v(w) = \frac{1}{C} \sum_{c=1}^C \frac{|v_{w,c} - v_{0,c}|}{v_{0,c} + \varepsilon} \quad (v_{0,c} = \sigma_{0,c}^2)$$

$$\sigma^2 \text{ drift (p95)} = \text{quantile}_{0.95}(\{d_v(w)\}_{w=1}^W)$$

3. Experiment

Data-Level Analysis: H6: 개별 수준 상관

- 개별 수준 상관이란, RevIN 이 mean / var 드리프트를 줄여서 그 실행에서 RMSE 도 좋아졌는지를 각 실행별로 볼 수 있는 지표
- Dataset D 와 Horizon H 를 고정하고 여러 Seed 에 대해 RevIN OFF – RevIN ON 의 차이 값을 구하여 피어슨 상관을 계산함
- 양의 상관은 드리프트를 줄일 수록 RMSE 개선도 있다는 의미, 무상관 또는 음의 상관은 드리프트 감소에 따른 RMSE 개선이 약하거나 나빠진다는 의미로 해석됨
- 각 Seed s 에 대해 계산은 다음과 같음

$$\Delta \text{RMSE}_s = \text{RMSE}_{OFF}(D, H, s) - \text{RMSE}_{ON}(D, H, s)$$

$$\Delta \mu_s^{abs} = \mu \text{ drift}_{OFF}^{abs}(D, H, s) - \mu \text{ drift}_{ON}^{abs}(D, H, s)$$

$$\Delta \sigma_s^{2abs} = \sigma^2 \text{ drift}_{OFF}^{abs}(D, H, s) - \sigma^2 \text{ drift}_{ON}^{abs}(D, H, s)$$

$$r = \text{corr}(\{\Delta \text{RMSE}_s\}, \{\Delta \mu_s^{abs}\}) \text{ or } r = \text{corr}(\{\Delta \text{RMSE}_s\}, \{\Delta \sigma_s^{2abs}\})$$

3. Experiment

Data-Level Analysis: H6

- H6 실험에 대해 Weather / ETT Datasets에 대한 데이터 자체 드리프트 분포 요약은 다음과 같음

Datasets	Horizon	μ drift (평균, SD)	μ drift (p95, SD)	σ^2 drift (평균, 상대)	σ^2 drift (p95, 상대)	Window 개수
ETTm2	96	1.1348	1.5944	0.8923	0.9518	290
	192	1.1310	1.5748	0.8724	0.9419	145
	336	1.1332	1.5507	0.8480	0.9308	82
	720	1.1280	1.4703	0.8083	0.9170	38
Datasets	Horizon	μ drift (평균, SD)	μ drift (p95, SD)	σ^2 drift (평균, 상대)	σ^2 drift (p95, 상대)	Window 개수
ETTh2	96	1.1343	1.5546	0.8454	0.9268	72
	192	1.1254	1.4597	0.8164	0.9164	36
	336	1.1336	1.4029	0.7818	0.8983	20
	720	1.1446	1.3548	0.7480	0.8531	9
Datasets	Horizon	μ drift (평균, SD)	μ drift (p95, SD)	σ^2 drift (평균, 상대)	σ^2 drift (p95, 상대)	Window 개수
Weather	96	0.5985	1.0940	0.9529	1.1229	219
	192	0.5763	1.0461	0.9185	1.1921	109
	336	0.5593	1.0011	0.8768	1.2435	62
	720	0.5364	1.0002	0.8374	1.2865	29

3. Experiment

Data-Level Analysis: H6

- H6 실험에 대해 Weather / ETT Datasets에 대한 데이터 자체 드리프트 분포 요약은 다음과 같음

Datasets	Horizon	μ drift (평균, SD)	μ drift (p95, SD)	σ^2 drift (평균, 상대)	σ^2 drift (p95, 상대)	Window 개수
ETTm1	96	0.5837	0.8377	0.8373	1.1900	290
	192	0.5715	0.7947	0.8116	1.0765	145
	336	0.5619	0.7660	0.7967	1.1183	82
	720	0.5390	0.7138	0.7383	0.9661	38

Datasets	Horizon	μ drift (평균, SD)	μ drift (p95, SD)	σ^2 drift (평균, 상대)	σ^2 drift (p95, 상대)	Window 개수
ETTh1	96	0.5546	0.7429	0.7828	1.0283	72
	192	0.5343	0.7427	0.7474	0.9606	36
	336	0.5285	0.6736	0.7169	0.8492	20
	720	0.5081	0.6079	0.6898	0.8045	9

- ETTm2, ETTh2에 대한 개별 수준 상관 표는 다음과 같음

Datasets	Horizon	Corr(Δ RMSE, $\Delta \mu$ drift (abs))	Corr(Δ RMSE, $\Delta \sigma^2$ drift (abs))
ETTm2	96	0.897	0.550
	192	0.822	0.185
	336	0.795	0.338
	720	0.715	0.427

Datasets	Horizon	Corr(Δ RMSE, $\Delta \mu$ drift (abs))	Corr(Δ RMSE, $\Delta \sigma^2$ drift (abs))
ETTm2	96	0.654	0.375
	192	-0.226	-0.367
	336	-0.593	-0.460
	720	-0.776	-0.741

3. Experiment

Data-Level Analysis: H6

- H6 실험에 대해 Weather / ETT Datasets에 따른 실험 결과는 다음과 같음

Datasets	Horizon	Model	Avg. RMSE (mean ± std)	Avg. MAE (mean ± std)	Δ RMSE (Improvements)
ETTm2	96	RevIN ON (Baseline)	2.96 ± 0.00	1.86 ± 0.00	—
		RevIN OFF	5.41 ± 1.34	3.90 ± 1.04	+ 45.31 %
	192	RevIN ON (Baseline)	3.34 ± 0.02	2.17 ± 0.01	—
		RevIN OFF	6.33 ± 0.74	4.61 ± 0.58	+ 47.34 %
	336	RevIN ON (Baseline)	3.69 ± 0.02	2.39 ± 0.01	—
		RevIN OFF	6.66 ± 0.12	4.85 ± 0.08	+ 44.60 %
	720	RevIN ON (Baseline)	4.10 ± 0.01	2.66 ± 0.00	—
		RevIN OFF	6.80 ± 0.10	4.94 ± 0.07	+ 39.66 %

Datasets	Horizon	Model	Avg. RMSE (mean ± std)	Avg. MAE (mean ± std)	Δ RMSE (Improvements)
ETTh2	96	RevIN ON (Baseline)	3.60 ± 0.01	2.37 ± 0.01	—
		RevIN OFF	5.96 ± 1.05	4.34 ± 0.84	+ 39.63 %
	192	RevIN ON (Baseline)	3.66 ± 0.02	2.42 ± 0.02	—
		RevIN OFF	5.92 ± 0.91	4.35 ± 0.73	+ 38.30 %
	336	RevIN ON (Baseline)	3.73 ± 0.02	2.46 ± 0.01	—
		RevIN OFF	6.05 ± 0.76	4.45 ± 0.61	+ 38.43 %
	720	RevIN ON (Baseline)	3.64 ± 0.01	2.40 ± 0.01	—
		RevIN OFF	6.42 ± 0.33	4.75 ± 0.26	+ 43.29 %

3. Experiment

Data-Level Analysis: H6

- H6 실험에 대해 Weather / ETT Datasets에 따른 실험 결과는 다음과 같음

Datasets	Horizon	Model	Avg. RMSE (mean ± std)	Avg. MAE (mean ± std)	Δ RMSE (Improvements)
ETTm1	96	RevIN ON (Baseline)	2.81 ± 0.01	1.53 ± 0.00	—
		RevIN OFF	2.97 ± 0.04	1.69 ± 0.03	+ 5.22 %
	192	RevIN ON (Baseline)	3.07 ± 0.01	1.69 ± 0.00	—
		RevIN OFF	3.22 ± 0.06	1.87 ± 0.05	+ 4.41 %
	336	RevIN ON (Baseline)	3.22 ± 0.00	1.82 ± 0.00	—
		RevIN OFF	3.47 ± 0.17	2.12 ± 0.15	+ 7.16 %
	720	RevIN ON (Baseline)	3.32 ± 0.01	1.95 ± 0.01	—
		RevIN OFF	3.67 ± 0.11	2.40 ± 0.09	+ 9.37 %

Datasets	Horizon	Model	Avg. RMSE (mean ± std)	Avg. MAE (mean ± std)	Δ RMSE (Improvements)
ETTh1	96	RevIN ON (Baseline)	3.05 ± 0.01	1.80 ± 0.01	—
		RevIN OFF	4.38 ± 0.06	2.81 ± 0.03	+ 30.28 %
	192	RevIN ON (Baseline)	3.18 ± 0.01	1.93 ± 0.01	—
		RevIN OFF	4.47 ± 0.05	2.88 ± 0.03	+ 28.94 %
	336	RevIN ON (Baseline)	3.30 ± 0.01	2.03 ± 0.01	—
		RevIN OFF	4.64 ± 0.07	3.00 ± 0.03	+ 28.96 %
	720	RevIN ON (Baseline)	3.15 ± 0.03	1.96 ± 0.02	—
		RevIN OFF	4.60 ± 0.05	3.07 ± 0.02	+ 31.47 %

3. Experiment

Data-Level Analysis: H6

- H6 실험에 대해 Weather / ETT Datasets 에 따른 실험 결과는 다음과 같음

Datasets	Horizon	Model	Avg. RMSE (mean \pm std)	Avg. MAE (mean \pm std)	Δ RMSE (Improvements)
Weather	96	RevIN ON (Baseline)	39.59 \pm 0.08	13.13 \pm 0.03	–
		RevIN OFF	51.12 \pm 0.42	20.19 \pm 0.22	+ 22.55 %
	192	RevIN ON (Baseline)	40.64 \pm 0.17	13.76 \pm 0.05	–
		RevIN OFF	52.38 \pm 1.26	20.89 \pm 0.49	+ 22.40 %
	336	RevIN ON (Baseline)	41.93 \pm 0.25	14.46 \pm 0.08	–
		RevIN OFF	54.02 \pm 1.98	21.78 \pm 0.54	+ 22.37 %
	720	RevIN ON (Baseline)	41.56 \pm 0.19	14.73 \pm 0.06	–
		RevIN OFF	56.09 \pm 1.78	22.55 \pm 0.58	+ 25.91 %