

Test-time State-augmented Adaptive Forecasting for Time Series Foundation Models

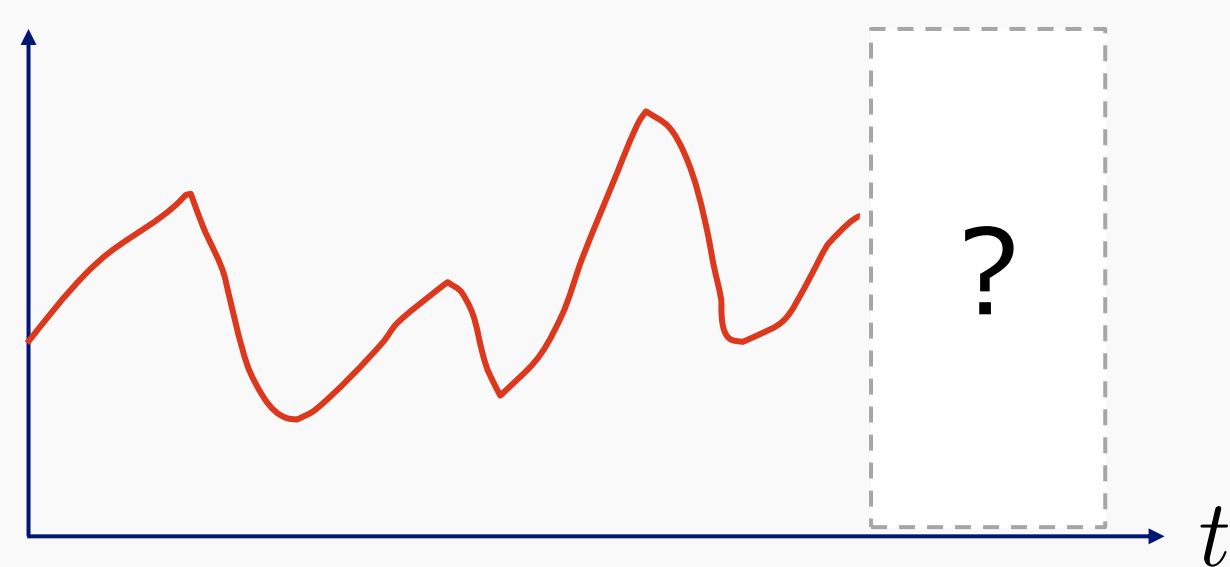
足立 一樹^{*,†}, 金井 関利^{*}, 山口 真弥^{*}, 熊谷 充敏^{*}, 濱上 知樹[†]

^{*}NTT株式会社 [†]横浜国立大学

時系列基盤モデル

時系列基盤モデル

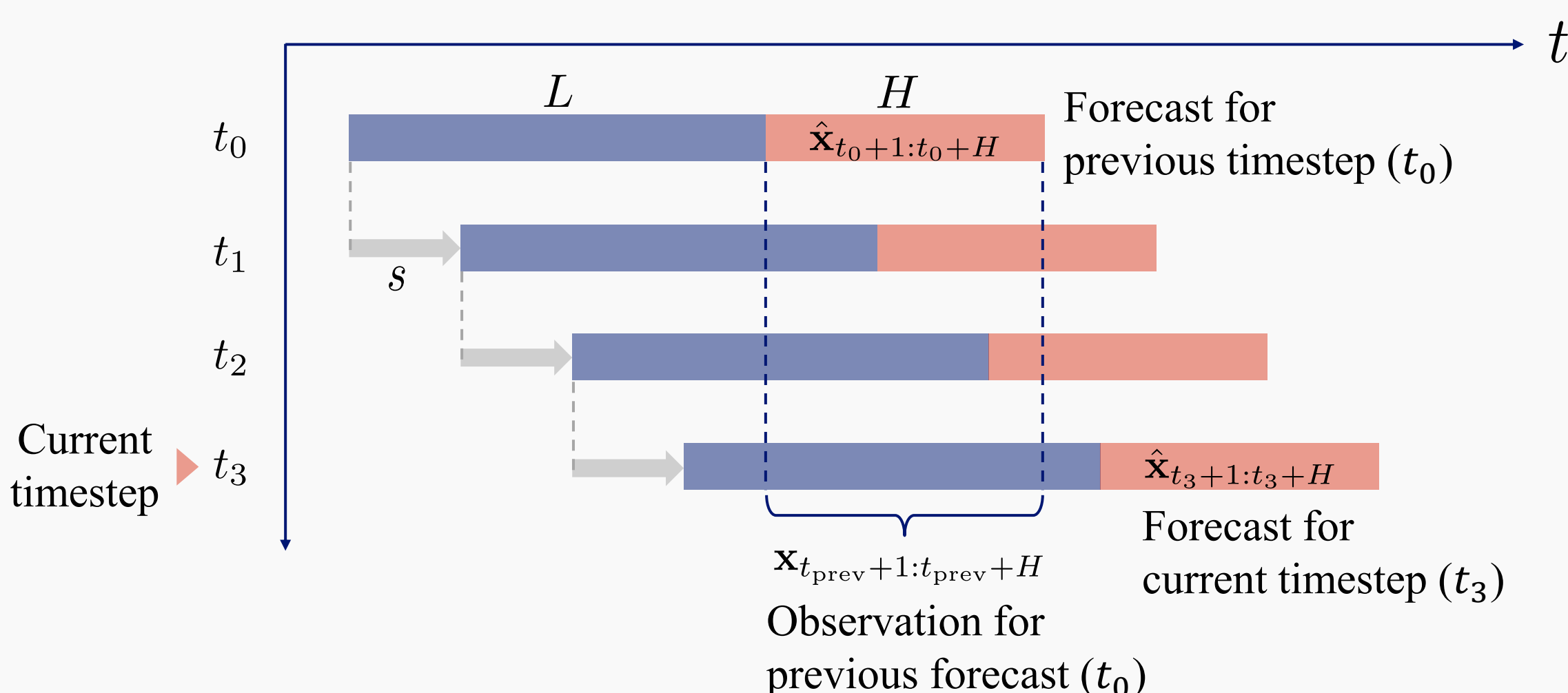
- LLMの成功 → 時系列予測でも同様のことができるか？
- 大量のデータで予測を学習させたモデル
- 初見の時系列データに対してゼロショット予測が可能



- 多くの学習済みモデルが公開
 - Lag-Llama (Morgan Stanley, 2023), TimesFM (Google, 2023), Chronos (AWS, 2024), Moirai (Salesforce, 2024), ...

時系列基盤モデルの推論

- Rolling window forecasting
 - 直近の固定長 L の系列（入力窓）から H 時刻先までを予測
 - 入力窓をずらしながら繰り返す



課題

計算量と精度のトレードオフ

- 多くの時系列基盤モデルはTransformer → 入力長 L の二乗オーダーで計算量が増加
- L を小さくすれば計算量を抑えられるが、精度が低下

Transformerは状態を持たない

- 時系列データは長期依存性を持つ場合があるが、入力窓より外側を考慮できない
- 入力が同じに見えても出力が違い得る（コンセプトドリフト）
 - L が小さいほど顕著に

目標：短い入力窓で、長期依存性や非定常性を捉えながら予測

問題設定

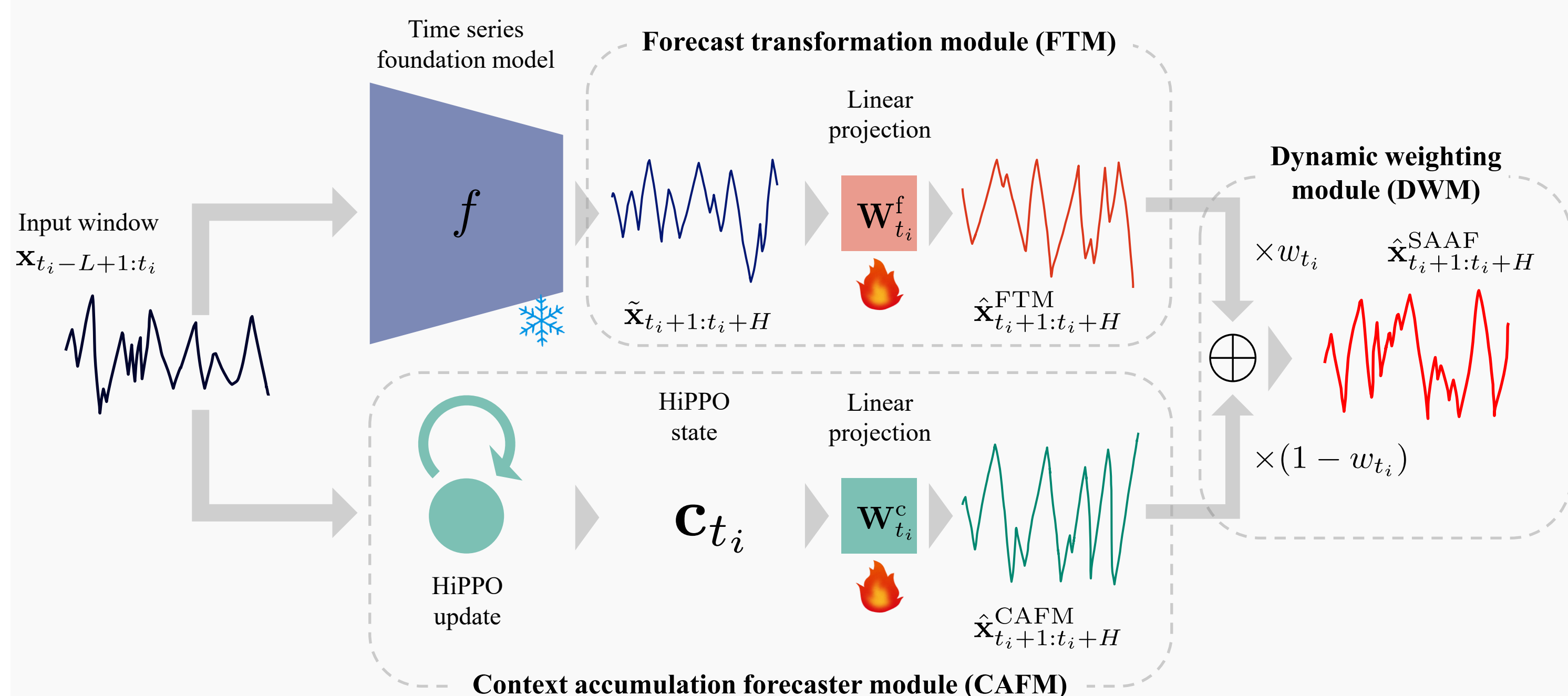
Input

- 長さ L の入力窓 $\{\mathbf{x}_{t_i-L+1:t_i}\}_{i=0}^T$ が時刻順に与えられる
- 各時刻 $t = si$ ごとに入力窓の H 時刻先までを基盤モデル f で予測

Output

- 時系列基盤モデルの予測を改善して出力

提案手法



長期依存性への対処 (CAFM)

- 入力系列を固定次元の状態ベクトル (HiPPO [1]) に蓄積
 - 入力系列を多項式基底に射影した係数 $\mathbf{c}_{t_i}$ がHiPPO
 - 逐次計算できる ($\bar{\mathbf{A}}, \bar{\mathbf{b}}$ は固定で学習不要)

$$\mathbf{c}_{t+1} = \bar{\mathbf{A}}\mathbf{c}_t + \bar{\mathbf{b}}x_t$$

- 実際の観測を用いてベイズ線形回帰で出力を学習

$$\hat{\mathbf{x}}_{t_i+1:t_i+H}^c = \mathbf{W}_{t_i}^c \mathbf{c}_{t_i}$$

非定常性への対処 (FTM)

- 直近の入出力関係の変化に対して適応
- 基盤モデルの予測を変換

$$\hat{\mathbf{x}}_{t_i+1:t_i+H}^f = \mathbf{W}_{t_i}^f f(\mathbf{x}_{t_i-L+1:t_i})$$

2つの予測のバランス

- 最近の誤差に基づき、正確な方の重みを増加

$$\hat{\mathbf{x}}_{t_i:t_i+H} = w_{t_i} \hat{\mathbf{x}}_{t_i:t_i+H}^c + (1 - w_{t_i}) \hat{\mathbf{x}}_{t_i:t_i+H}^f$$

$$w_{t_i} = \text{sigmoid}(\gamma(\bar{e}_{t_i}^f - \bar{e}_{t_i}^c))$$

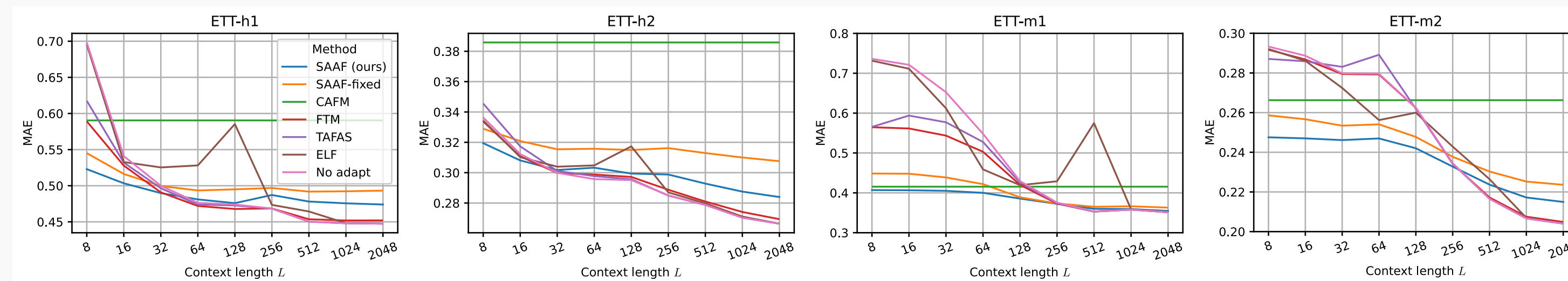
$$\bar{e}_{t_i}^{\{c,f\}} = m\bar{e}_{t_{i-1}}^{\{c,f\}} + (1 - m) \text{MSE}(\hat{\mathbf{x}}_{t_{\text{prev}}:t_{\text{prev}}+H}^{\{c,f\}}, \mathbf{x}_{t_{\text{prev}}:t_{\text{prev}}+H})$$

[1] Gu et al., HiPPO: Recurrent Memory with Optimal Polynomial Projections, NeurIPS2020.

実験

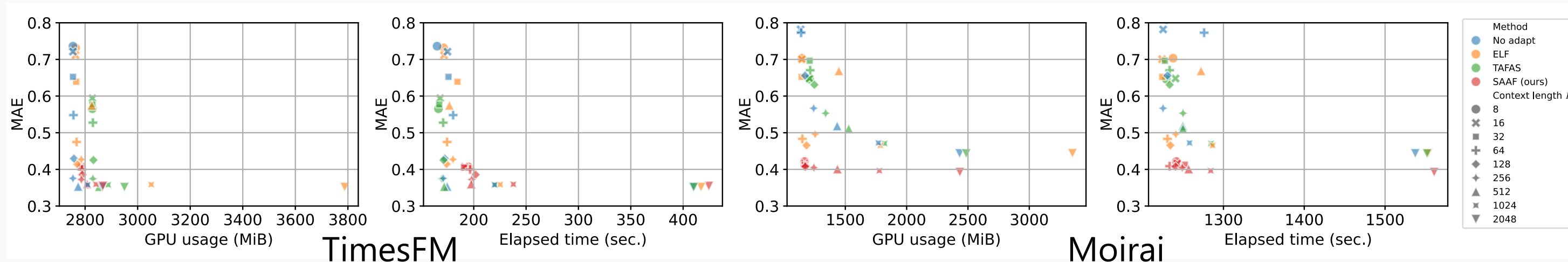
予測精度

- 短いコンテキスト長でも精度低下を抑制 (TimesFM)



計算コスト

- 推論が重いモデルで効果大



予測の可視化

- CAFMは初期にWarmupが必要だが相補的に予測

