AdaPTS: Adapting Univariate Foundation Models to Probabilistic Multivariate Time Series Forecasting

Abdelhakim Benechehab[1,2], Vasilii Feofanov[1], Giuseppe Paolo[1], Albert Thomas[1], Maurizio Filippone[3], Balázs Kégl[1]

[1] Huawei Noah's Ark Lab, Paris, France
[2] Department of Data Science, EURECOM
[3] Statistics Program, KAUST

-

June 15, 2025

Outline

Problem setup

Consider a multivariate time series forecasting task:

- $\mathbf{X} \in \mathbb{R}^{L \times D}$ data matrix
- $\mathbf{Y} \in \mathbb{R}^{H \times D}$ target
  - L lookback window (context length)
  - H forecasting horizon
  - D dimension (number of covariates)

We want to find the best adapter φ^* such that:

Definition (adapter)

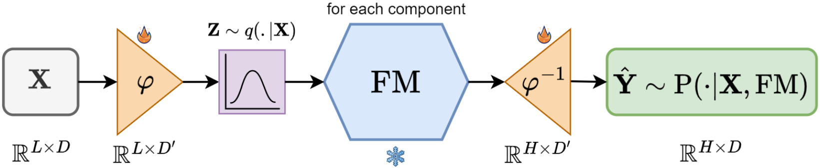
Feature-space transformation $\varphi : \mathbb{R}^D \to \mathbb{R}^{D'}$ such that:

$$\hat{\mathbf{Y}}(\mathbf{X}; \varphi) = \varphi^{-1}\big(\texttt{FM}(\varphi(\mathbf{X}))\big), \text{ and } \varphi^* = \text{argmin}_{\varphi} \|\mathbf{Y} - \hat{\mathbf{Y}}(\mathbf{X}; \varphi)\|_{\mathrm{F}}^2,$$

where FM is a fixed time series foundation model.

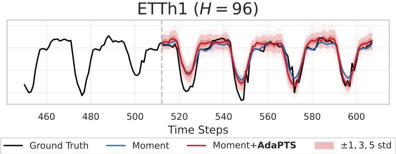
Outline

AdaPTS: Adapters for Probabilistic multivariate Time Series forecasting [1]



Properties:

1. Mixing features
2. Probabilistic predictions



Outline

The linear case

For a linear adapter $\varphi(\mathbf{X}) = \mathbf{X}\mathbf{W}_\varphi$ and a linear FM $f_{\mathrm{FM}}(\mathbf{X}) = \mathbf{W}_{\mathsf{FM}}^\top \mathbf{X} + \mathbf{b}_{\mathsf{FM}} \mathbf{1}^\top$, we have:

Proposition (Optimal linear adapter)

The closed-form solution of the problem

$$\min_{\mathbf{W}_\varphi} \mathcal{L}(\mathbf{W}_\varphi) = \|\mathbf{Y} - \left(\mathbf{W}_{FM}^\top \mathbf{X}\mathbf{W}_\varphi + \mathbf{b}_{FM}\mathbf{1}^\top\right)\mathbf{W}_\varphi^{-1}\|_F^2$$

writes as:

$$\mathbf{W}_\varphi^* = (\mathbf{B}^\top \mathbf{A})^+ \mathbf{B}^\top \mathbf{B},$$

where $\mathbf{A} = \mathbf{Y} - \mathbf{W}_{FM}^\top \mathbf{X}$ *and* $\mathbf{B} = \mathbf{b}_{FM}\mathbf{1}^\top$.

$\rightarrow$ Takeaway: The optimal solution is **not** the identity.

Families of adapters

1. deterministic
   - Linear AutoEncoder
   - Deep non-linear AutoEncoder
   - Normalizing Flow
2. probabilistic
   - $+$ Variational Inference
   - $+$ MC Dropout

ELBO-like training objective for variational adapters:

Proposition

Maximization of an ELBO-like lower bound on the marginal likelihood of the target $\mathbf{Y}$:

$$\log p_\theta(\mathbf{Y}|\mathbf{X}, f_{\mathrm{FM}}) \geq \mathbb{E}_{q_\phi(\mathbf{Z}|\mathbf{X})}\left[\log p_\theta(\mathbf{Y}|\mathbf{X}, f_{\mathrm{FM}}(\mathbf{Z}))\right] - \mathrm{KL}\left(q_\phi(\mathbf{Z}|\mathbf{X}) \,\|\, p(\mathbf{Z})\right),$$

where KL denotes the Kullback-Leibler divergence.

Outline

Results: Forecasting error (MSE)

Benchmarking using `Moment` FM:

Dataset	H	No adpt	with adapter				
		`Moment`	`PCA`	`LinAE`	`dropLAE`	`LinVAE`	`VAE`
ETTh1	96	$0.411_{\pm.012}$	$0.433_{\pm.001}$	$0.402_{\pm.002}$	$\mathbf{0.395_{\pm.003}}$	$0.400_{\pm.001}$	$0.404_{\pm.001}$
	192	$\mathbf{0.431_{\pm.001}}$	$0.440_{\pm.000}$	$0.452_{\pm.002}$	$0.446_{\pm.001}$	$0.448_{\pm.002}$	$\mathbf{0.431_{\pm.001}}$
Ill	24	$2.902_{\pm.023}$	$2.98_{\pm.001}$	$2.624_{\pm.035}$	$2.76_{\pm.061}$	$2.542_{\pm.036}$	$\mathbf{2.461_{\pm.008}}$
	60	$3.000_{\pm.004}$	$3.079_{\pm.000}$	$3.110_{\pm.127}$	$2.794_{\pm.015}$	$\mathbf{2.752_{\pm.040}}$	$2.960_{\pm.092}$
Wth	96	$0.177_{\pm.010}$	$0.176_{\pm.000}$	$0.169_{\pm.000}$	$\mathbf{0.156_{\pm.001}}$	$0.161_{\pm.001}$	$0.187_{\pm.001}$
	192	$0.202_{\pm.000}$	$0.208_{\pm.001}$	$\mathbf{0.198_{\pm.001}}$	$0.200_{\pm.001}$	$0.204_{\pm.000}$	$0.226_{\pm.000}$
ExR	96	$\mathbf{0.130_{\pm.011}}$	$0.147_{\pm.000}$	$0.167_{\pm.013}$	$\mathbf{0.130_{\pm.011}}$	$0.243_{\pm.039}$	$0.455_{\pm.010}$
	192	$\mathbf{0.210_{\pm.002}}$	$0.222_{\pm.000}$	$0.304_{\pm.005}$	$0.305_{\pm.013}$	$0.457_{\pm.020}$	$0.607_{\pm.021}$

Validating other FMs on the Illness task:

FM	**Optimal adapter**	**MSE Imp (%)**
`TTM`	`VAE` $(\beta = 2, \sigma = 1)$	$12.35_{\pm 2.51}$
`TimesFM`	`dropoutLAE` $(p = 0.1)$	$3.64_{\pm 2.81}$

Desirable representation learning properties:

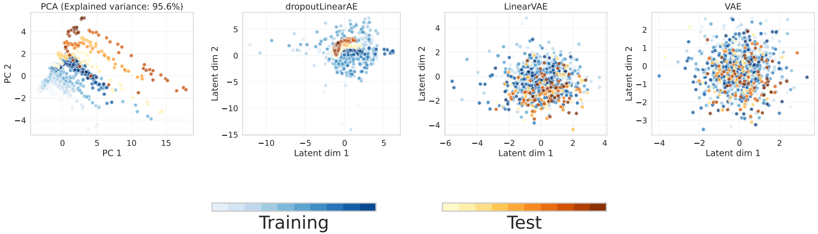
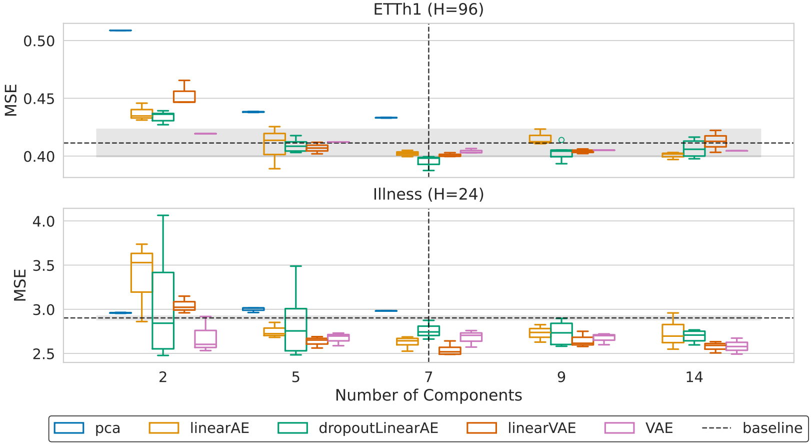


Figure: Visualization of the latent representation obtained by different adapters on Illness($H = 24$). Shaded colors indicate the time dimension, with lighter colors representing earlier timesteps.

Results: Dimensionality reduction

Better forecasting accuracy even with lower dimensions



Outline

Conclusion

- We presented **AdaPTS** a learning-based and probabilistic framework for adapting FMs to multivariate time series forecasting

Take Home Message

Foundation Models are powerful predictors trained on vast amounts of data
$\rightarrow$ **Adapters** are an effective way to adapt them to custom problems

References

A. Benechehab, V. Feofanov, G. Paolo, A. Thomas, M. Filippone, and B. Kégl, "Adapts: Adapting univariate foundation models to probabilistic multivariate time series forecasting," *Forty-second International Conference on Machine Learning (ICML)*, May 2025.

Thank You!

Want to know more?

Contact:

https://abenechehab.github.io/

abdelhakim.benechehab@gmail.com

@abenechehab

Slides available at:

https://abenechehab.github.io/assets/pdf/adapts.pdf