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

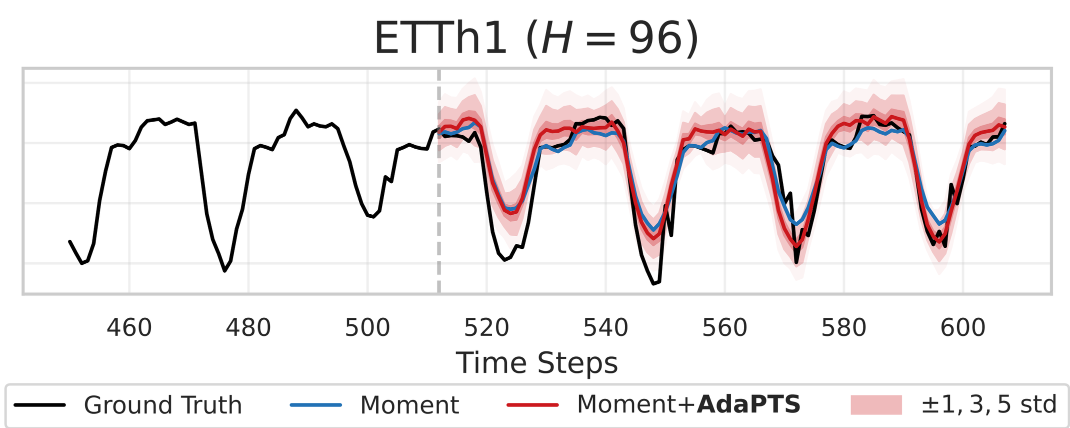
Abdelhakim Benechehab [1,2], Vasilii Feofanov[1], Giuseppe Paolo[1], Albert Thomas[1], Maurizio Filippone[3], Balázs Kégl[1]

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

ICLR

Motivation

- Many forecasting Foundation Models (*e.g.,* **Moment**) are univariate and deterministic
- How can we make them **Multivariate** and **Probabilistic** without changing their weights?



Problem Setup

- *Input:* multivariate time series $\mathbf{X} \in \mathbb{R}^{L \times D}$.
- *Output:* $\mathbf{Y} \in \mathbb{R}^{H \times D}$ with H being the forecasting horizon.
- *What we want:* use a pre-trained foundation model to predict $\mathbf{Y}$: $f_{\text{FM}}(\mathbf{X})$

Adapter. A *feature space* transformation $\varphi : \mathbb{R}^D \to \mathbb{R}^{D'}$ such that:

$$\hat{\mathbf{Y}}(\mathbf{X}; \varphi) = \varphi^{-1}\big(f_{\text{FM}}(\varphi(\mathbf{X}))\big)$$

<u>Goal:</u> find the optimal adapter:

$$\arg\min_{\varphi} \|\mathbf{Y} - \varphi^{-1}\big(f_{\text{FM}}(\varphi(\mathbf{X}))\big)\|_{\text{F}}^2$$

Analysing the linear case

Linear parametrization of the adapter: $\varphi(\mathbf{X}) = \mathbf{X}\mathbf{W}_\varphi$ where $\mathbf{W}_\varphi \in \mathbb{R}^{D \times D}$.

Assumptions:

- $\mathbf{W}_\varphi$ is full rank.
- Linear predictor: $f_{\text{FM}}(\mathbf{X}) = \mathbf{W}_{\text{FM}}^\top \mathbf{X} + \mathbf{b}_{\text{FM}} \mathbf{1}^\top$.

Proposition. the solution of:

$$\min_{\mathbf{W}_\varphi} \|\mathbf{Y} - (\mathbf{W}_{\text{FM}}^\top \mathbf{X}\mathbf{W}_\varphi + \mathbf{b}_{\text{FM}}\mathbf{1}^\top)\mathbf{W}_\varphi^{-1}\|_F^2$$

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 $\mathbf{B} = \mathbf{b}_{FM}\mathbf{1}^\top$.

insight: The optimal adapter is **NOT** the identity!

Probabilistic Adapters

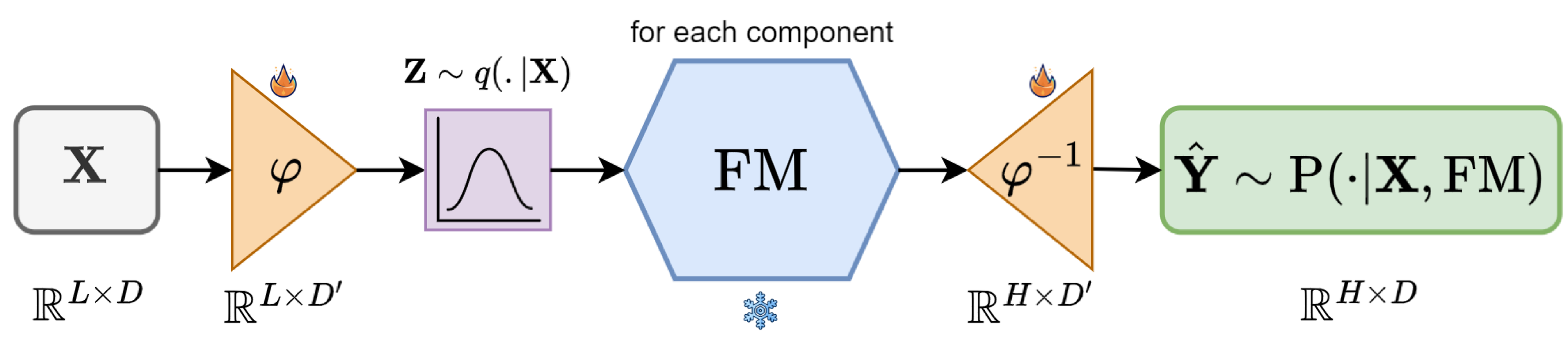
In practice, we learn an *encoder* $\varphi \triangleq \text{enc}_\phi$ and *decoder* $\varphi^{-1} \triangleq \text{dec}_\theta$.

Using Variational Inference, we can build <u>Probabilistic</u> forecasts:

Proposition. VAE training objective:

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

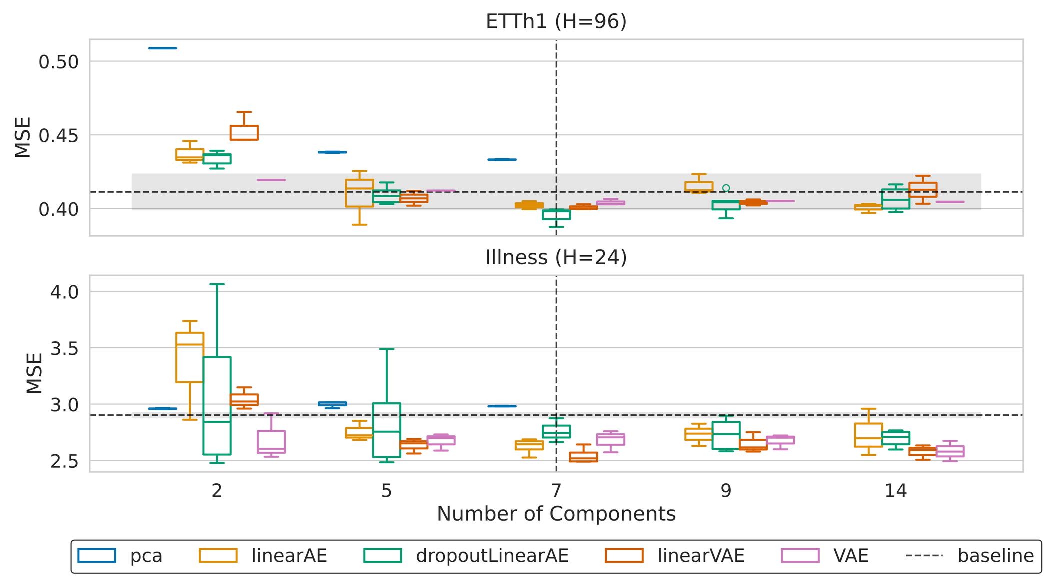
Our approach: AdaPTS



Forecasting error - MSE

Dataset	H	No adapter	with adapter				
		Moment	PCA	LinearAE	dropoutLAE	LinearVAE	VAE
ETTh1	96	$0.411_{\pm 0.012}$	$0.433_{\pm 0.001}$	$0.402_{\pm 0.002}$	$\mathbf{0.395_{\pm 0.003}}$	$0.400_{\pm 0.001}$	$0.404_{\pm 0.001}$
	192	$\mathbf{0.431_{\pm 0.001}}$	$0.440_{\pm 0.000}$	$0.452_{\pm 0.002}$	$0.446_{\pm 0.001}$	$0.448_{\pm 0.002}$	$\mathbf{0.431_{\pm 0.001}}$
Illness	24	$2.902_{\pm 0.023}$	$2.98_{\pm 0.001}$	$2.624_{\pm 0.035}$	$2.76_{\pm 0.061}$	$2.542_{\pm 0.036}$	$\mathbf{2.461_{\pm 0.008}}$
	60	$3.000_{\pm 0.004}$	$3.079_{\pm 0.000}$	$3.110_{\pm 0.127}$	$2.794_{\pm 0.015}$	$\mathbf{2.752_{\pm 0.040}}$	$2.960_{\pm 0.092}$
Weather	96	$0.177_{\pm 0.010}$	$0.176_{\pm 0.000}$	$0.169_{\pm 0.000}$	$\mathbf{0.156_{\pm 0.001}}$	$0.161_{\pm 0.001}$	$0.187_{\pm 0.001}$
	192	$0.202_{\pm 0.000}$	$0.208_{\pm 0.001}$	$\mathbf{0.198_{\pm 0.001}}$	$0.200_{\pm 0.001}$	$0.204_{\pm 0.000}$	$0.226_{\pm 0.000}$

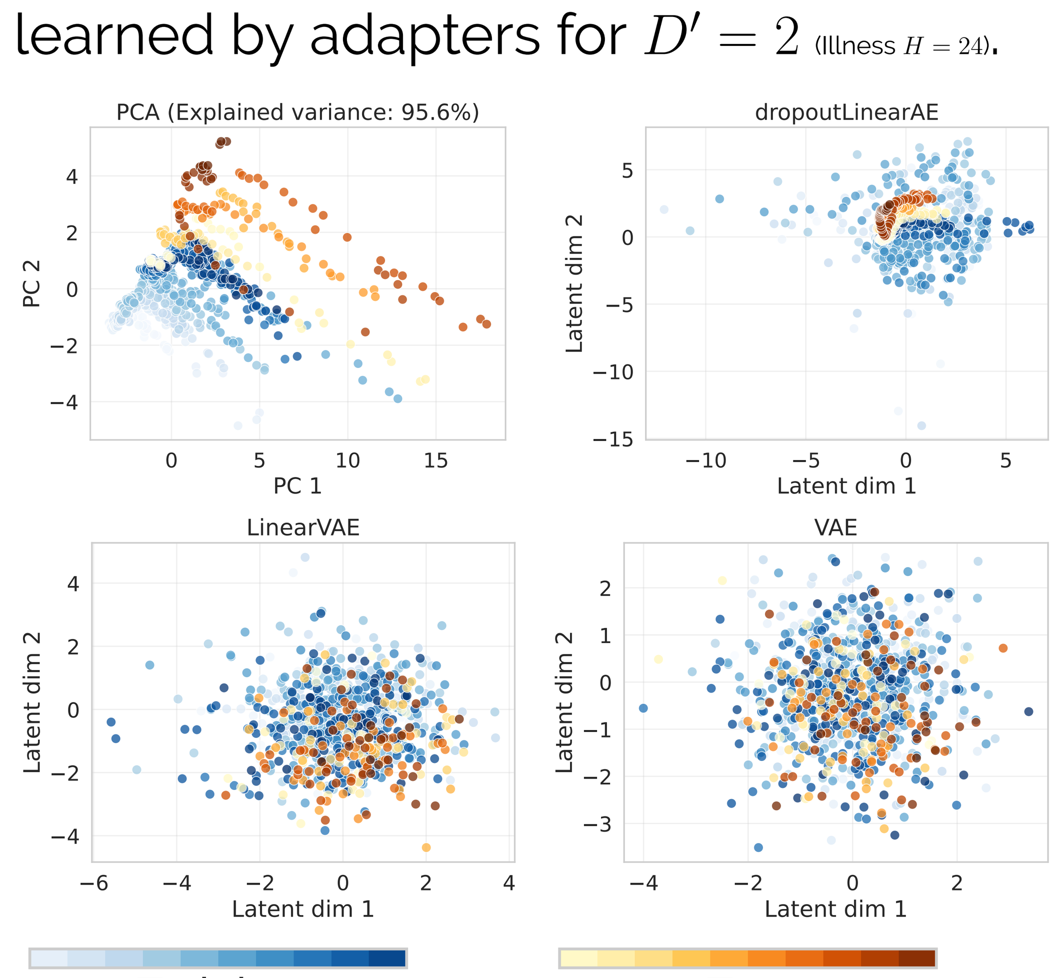
Dimensionality Reduction



► VAE adapter achieve better or similar performance at only $D' = 2$.

Latent representation

We visualize the latent representation learned by adapters for $D' = 2$ (Illness $H = 24$).



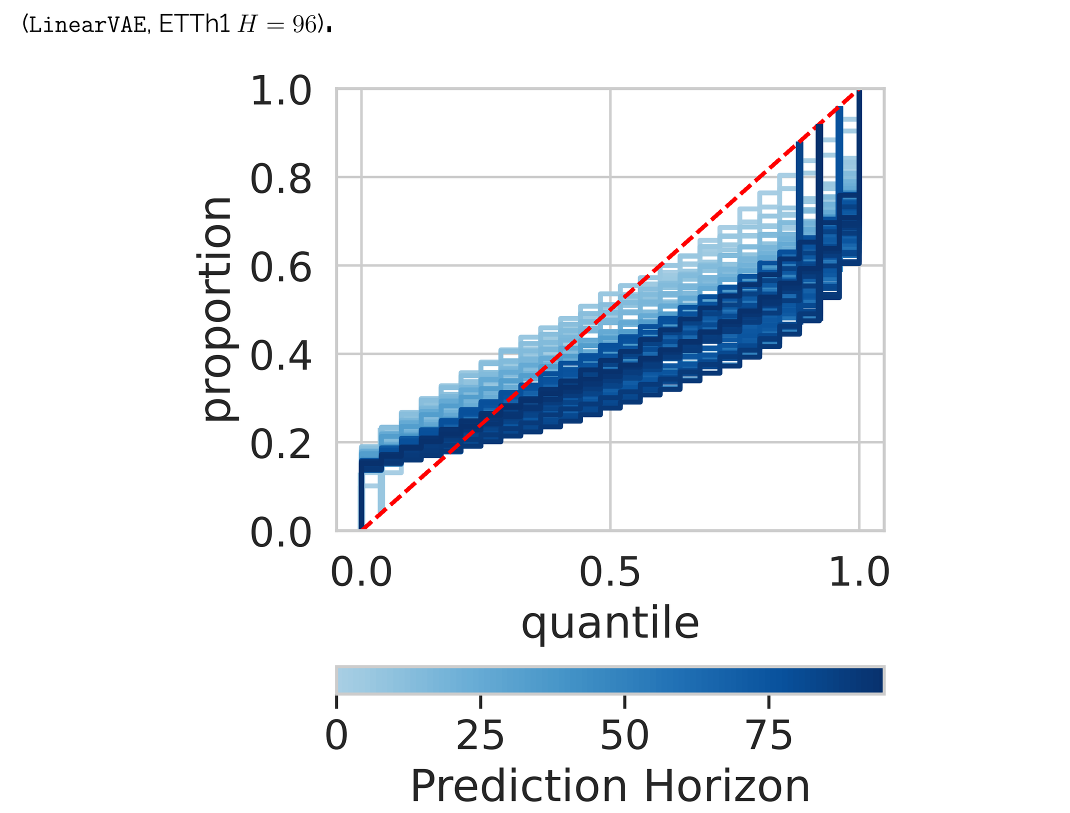
► Adapters help reduce **distribution shift** between training and test data.

Choice of adapters

Adapter	$D' < D$	Prob	non-linear
Identity	✗	✗	✗
PCA	✓	✗	✗
Dropout LinAE	✓	✓	✗
VAE	✓	✓	✓

Calibration

Quantile calibration reliability diagram



Take Home Message

AdaPTS is a **lightweight** solution that enables **channel-mixing**, **probabilistic forecasting**, and **dimensionality reduction** when using univariate TSFMs.

Main References

- **Benechehab et al.** - ICLR 2025
  Zero-shot Model-based Reinforcement Learning using Large Language Models
- **Ilbert, Feofanov et al.** - ICDEW 2025
  User-friendly Foundation Model Adapters for Multivariate Time Series Classification

Want to Know More?

paper & code author

abdelhakim.benechehab@gmail.com