

Are Time-Indexed Foundation Models the Future of Time Series Imputation?

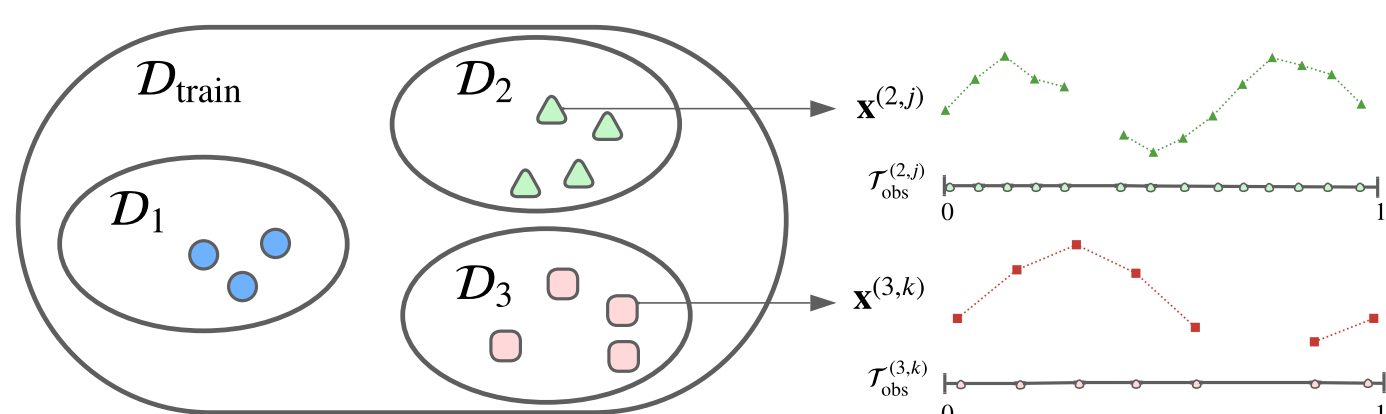
Etienne Le Naour^{1*} Tahar Nabil^{1*} Adrien Petralia¹ Ghislain Agoua¹

¹EDF R&D, Palaiseau, France

* Equal contribution

Context

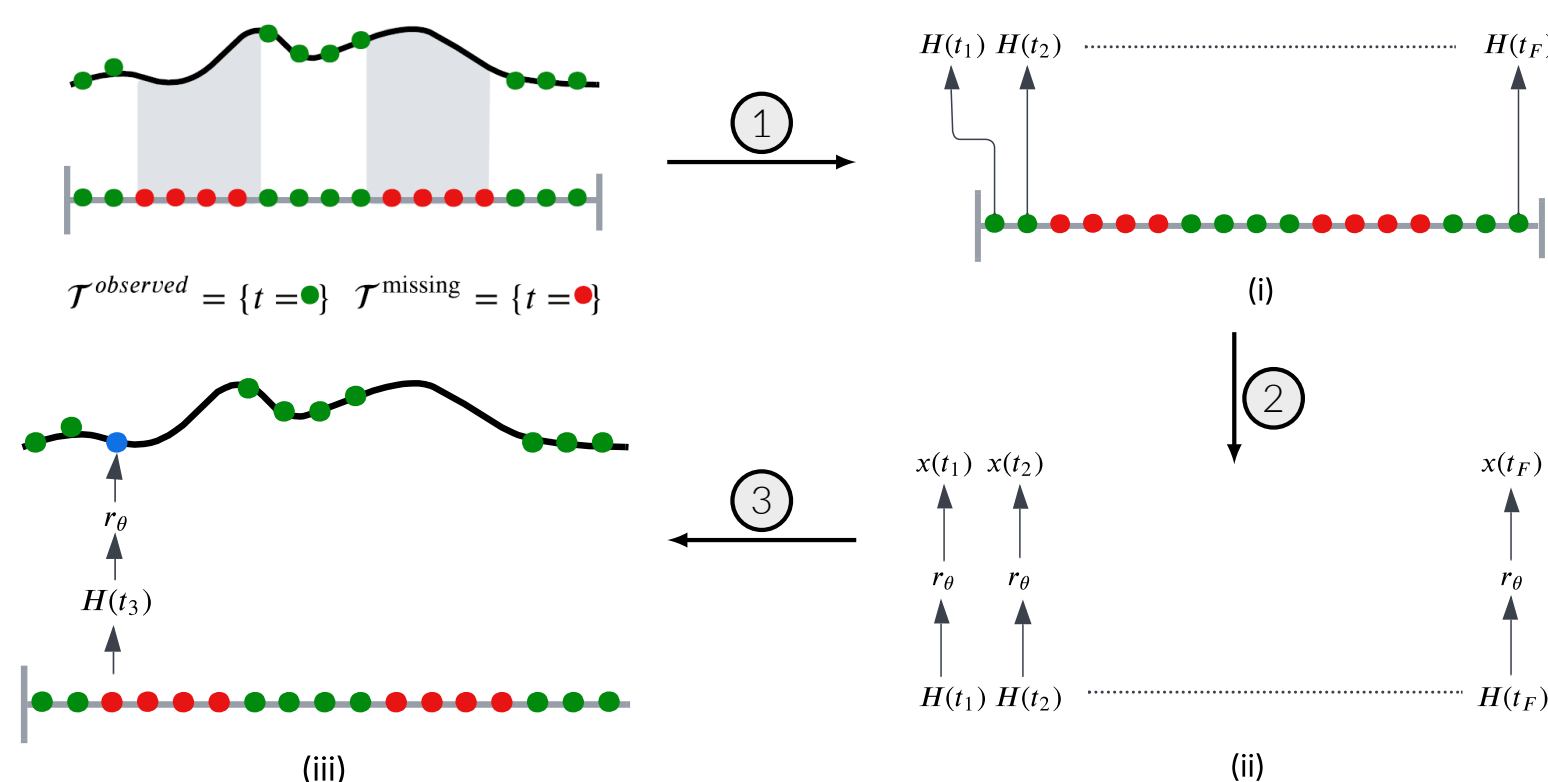
- Time Series Foundation models are a promise of **flexible** and **high-performing** models that compensate for the lack of data, withstand regime shifts, and democratize the use of advanced methods.
- These models have been widely studied for forecasting purposes, but they have been **much less explored for imputation**.
- Patch-based models are **not inherently designed** to handle the diverse missingness patterns of time series, which may be unaligned, irregular, and collected at varying sampling rates.



Time-Indexed Foundation Models

These models operate in three stages: (i) **encoding time**; (ii) **solving the regression problem** based on the observed context; and (iii) **inferring the values**.

- TabPFN-TS: handcrafted time features + deep regressor pretrained on synthetic toy problems with *in-context learning* [1]
- MoTM: deep adaptive features of time + ridge regressor [2]



Large-scale Univariate Benchmark

- 33 Out-of-Domain univariate datasets (climate, energy, traffic, etc.) with sampling freq from 5min to 1h
- 4 imputation settings: (i) 50% or (ii) 70% of observations missing at random, (iii) 2 or (iv) 4 one-day missing blocks
- 1.3M windows to impute in total**
- 6 supervised baselines: BRITS, CSDI, SAITS, TimesNet, TimeMixer++, TSLANet
- 3 local baselines: Linear Interpolation, Seasonal Naive, LOCF

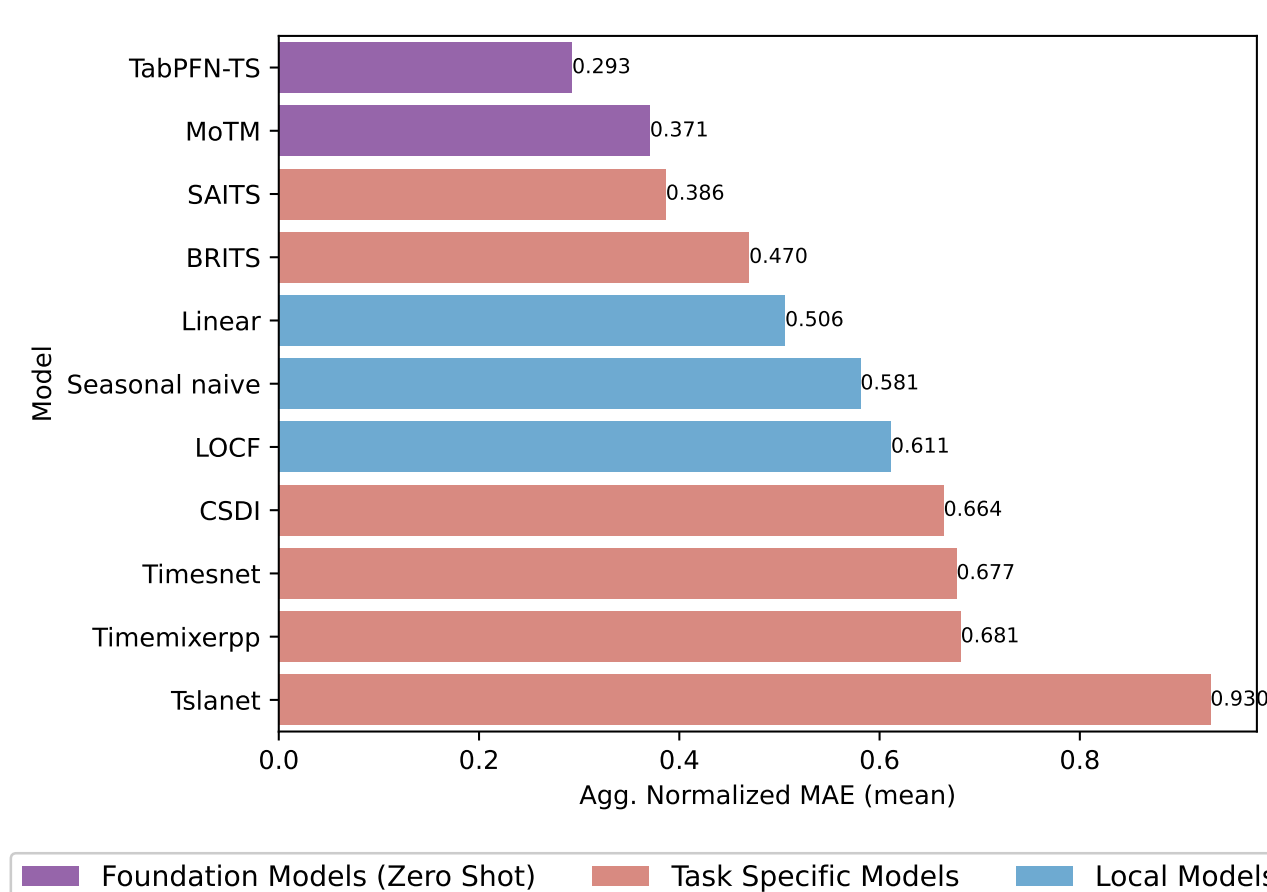


Figure 1. Univariate Benchmark (z-normalized MAEs) on 33 Out-of-Domain datasets.

- Time-index foundation models lead the benchmark
- Task-specific models show limited robustness
- Local baselines remain resilient

Uncertainty Quantification

TabPFN-TS & MoTM natively support quantile estimation: experiments on 11 datasets show the superiority of TabPFN-TS and limited generalization abilities of supervised baselines CSDI & SAITS.

Table 1. Weighted Quantile Loss (WQL) average scores on eleven representative univariate datasets.

	TabPFN-TS	MoTM	CSDI	SAITS-Q
WQL	0.241	<u>0.316</u>	0.453	0.476

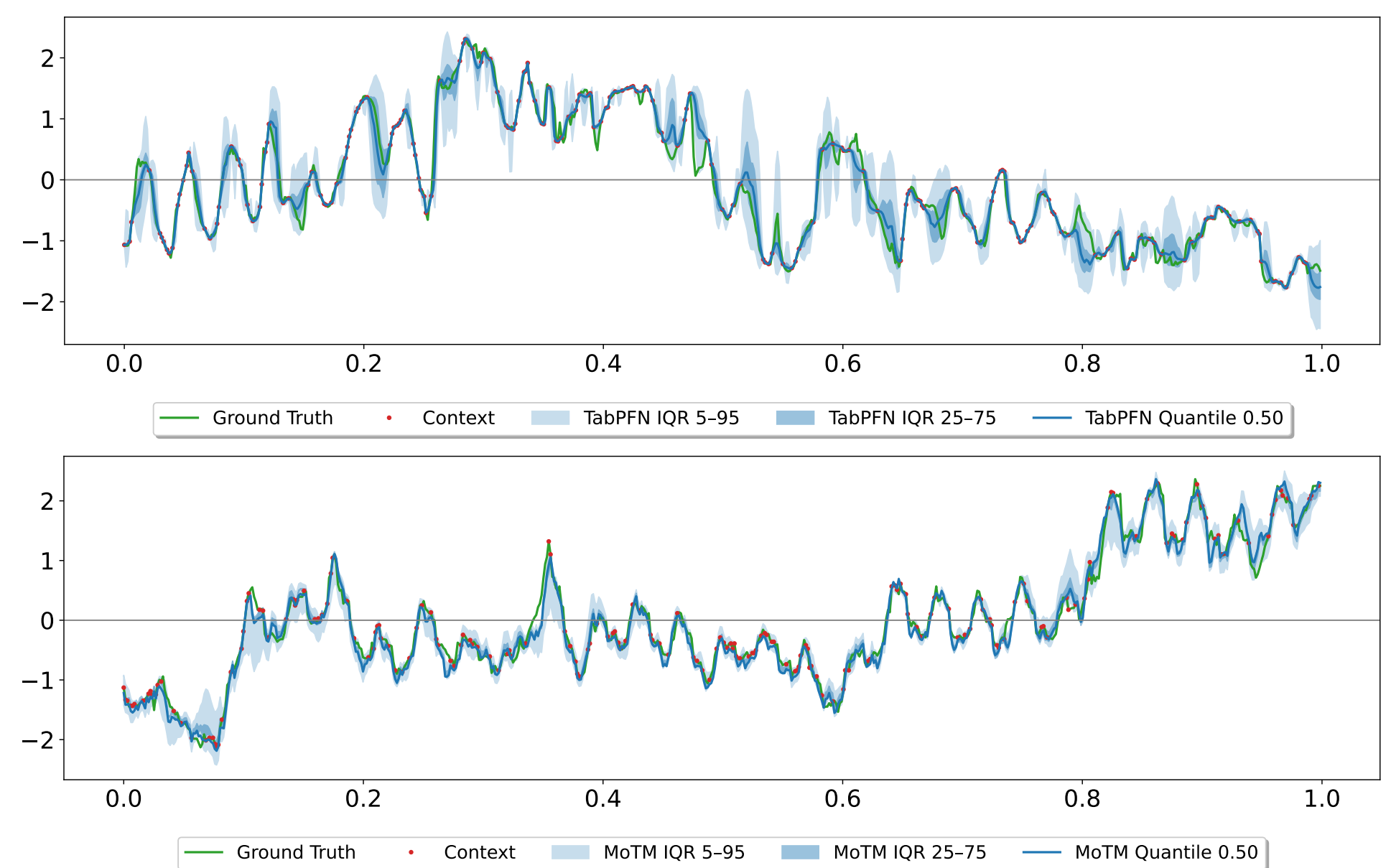


Figure 2. Qualitative quantile results for the 70% missing values scenario, resp. TabPFN-TS and MoTM.

Zero-shot Integration of Covariates

- Additional covariates $x(t)$ integrated at test-time by simple concatenation to the time features $H(t) \rightarrow$ **no finetuning**
- Yields substantial gains when the covariate strongly informs about the target

Table 2. Complete MAE results on datasets with covariates.

Dataset	Setting	TabPFN-TS		MoTM		Other baselines	
		TabPFN-TS (W/ Covar)	TabPFN-TS (W/o Covar)	MoTM (W/ Covar)	MoTM (W/o Covar)	Ridge on Covar	SAITS Multivar
PV-France	Pointwise 1	0.045	0.109	0.054	0.102	0.115	0.390
	Pointwise 2	0.051	0.160	<u>0.069</u>	0.131	0.123	0.443
	Blocks 1	0.052	0.175	<u>0.086</u>	0.191	0.104	0.496
	Blocks 2	0.049	0.190	<u>0.086</u>	0.179	0.106	0.485
Wind-France	Pointwise 1	0.101	0.098	0.128	0.186	0.318	0.359
	Pointwise 2	0.138	0.153	0.161	0.244	0.322	0.408
	Blocks 1	0.275	0.470	0.335	0.600	<u>0.321</u>	0.590
	Blocks 2	0.248	0.470	0.317	0.594	<u>0.335</u>	0.581
Load-France	Pointwise 1	0.037	0.037	0.138	0.146	0.667	0.292
	Pointwise 2	0.056	<u>0.059</u>	0.158	0.164	0.667	0.321
	Blocks 1	0.143	<u>0.146</u>	0.236	0.243	0.669	0.490
	Blocks 2	0.170	<u>0.178</u>	0.262	0.270	0.674	0.498
W. Covariate improvement		/	31.54%	/	31.03%	/	/

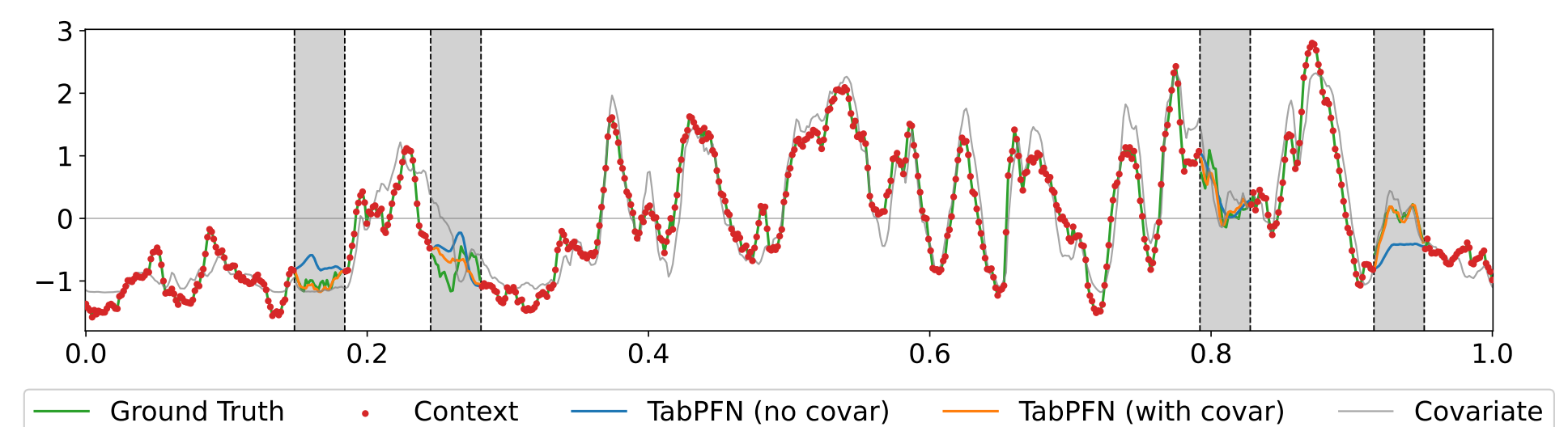


Figure 3. Wind-France dataset. TabPFN-TS qualitative results W. and W/o covariates.



Link to
Paper

- Shi Bin Hoo, Samuel Müller, David Salinas, and Frank Hutter. From tables to time: How tabpfn-v2 outperforms specialized time series forecasting models. *arXiv preprint arXiv:2501.02945*, 2025.
- Etienne Le Naour, Tahar Nabil, and Ghislain Agoua. MoTM: Towards a foundation model for time series imputation based on continuous modeling. *ECML / PKDD 2025 AALTD Workshop*, 2025.