MoTM: Towards a Foundation Model for Time Series Imputation based on Continuous Modeling

Workshop AALTD, ECML / PKDD 2025

September 19, 2025

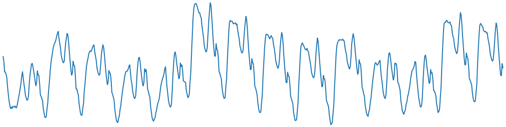
Etienne Le Naour * etienne.le-naour@edf.fr

Tahar Nabil * tahar.nabil@edf.fr

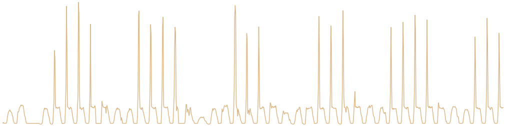
Ghislain Agoua ghislain.agoua@edf.fr

* equal contribution

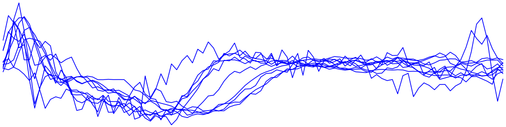
The Traditional Paradigm: One Model per Dataset



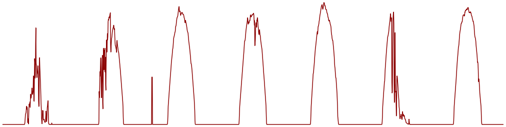
Learn f_{θ_1} on aggregated load curves



Learn f_{θ_3} on road occupancies

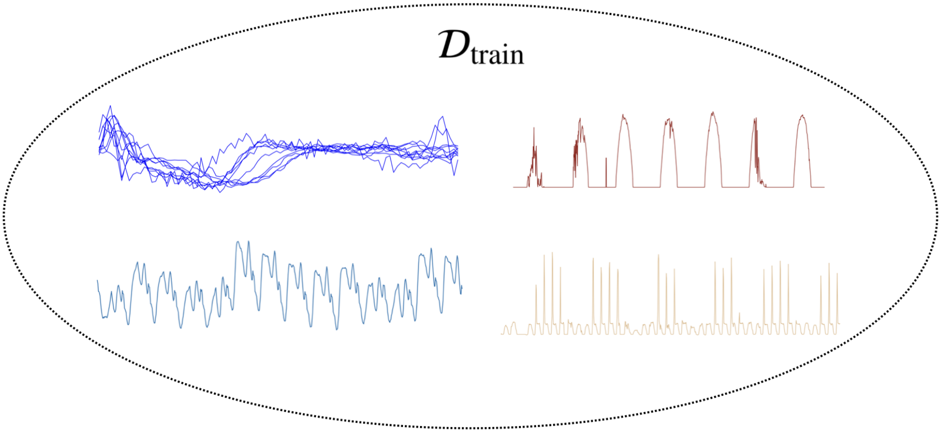


Learn f_{θ_2} on heartbeat electrical activity



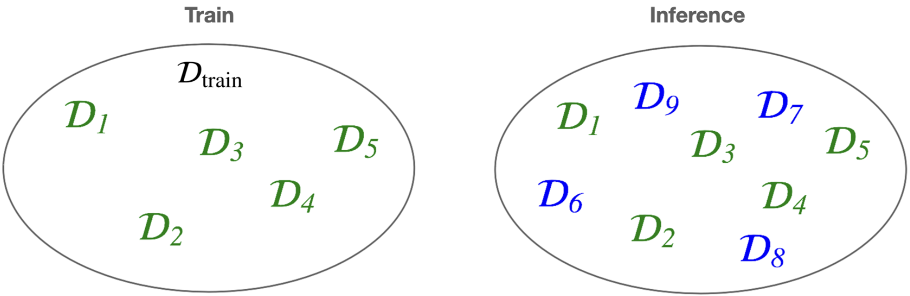
Learn f_{θ_4} on solar power generation

A New Paradigm with Foundation Models: Joint Training on Multiple Datasets



Learn a single model f_θ on a collection of training datasets

Generalization at Inference



At inference time, we aim for two generalizations:

(i) *In-Domain:* perform well for new time series from the training datasets ($\mathcal{D} \subset \mathcal{D}_{\text{train}}$).

(ii) *Out-of-Domain (OOD):* perform well for entirely new datasets not seen during training ($\mathcal{D}_{\text{new}} \not\subset \mathcal{D}_{\text{train}}$) (in a 0-shot manner).

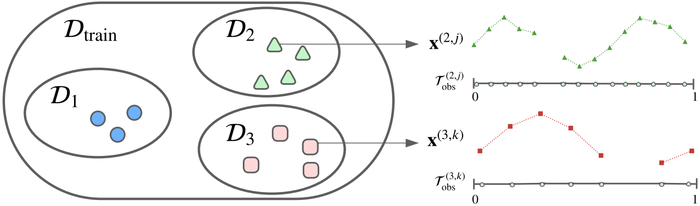
Extensively Studied in Forecasting, Limited in Imputation

Foundation models for forecasting

- Moirai [9], TimesFM [4], Chronos [1], ...

Foundation models for imputation are less explored

- NuwaTS [3], Moment [6]
- Learning jointly from multiple TS datasets presents several challenges, incl.: (i) varying sampling rates, (ii) unaligned timestamps



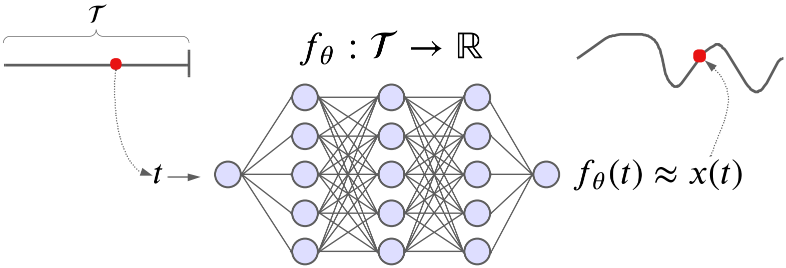
Introducing MoTM

- We propose **MoTM (Mixture of TimeFlow Models)**, a mixture of continuous neural representations (implicit neural representations) designed as a foundation model for time series imputation.

Method	**Handles Various Missing Patterns**	**Supports Multiple Sampling Rates**	**Performs OOD Inference**
BRITS [2], SAITS [5]	✓	✗	✗
NuwaTS [3], MOMENT [6]	✓	✗	✓
TimeFlow [7], ImputeINR [8]	✓	✓	✗
MoTM (Ours)	✓	✓	✓

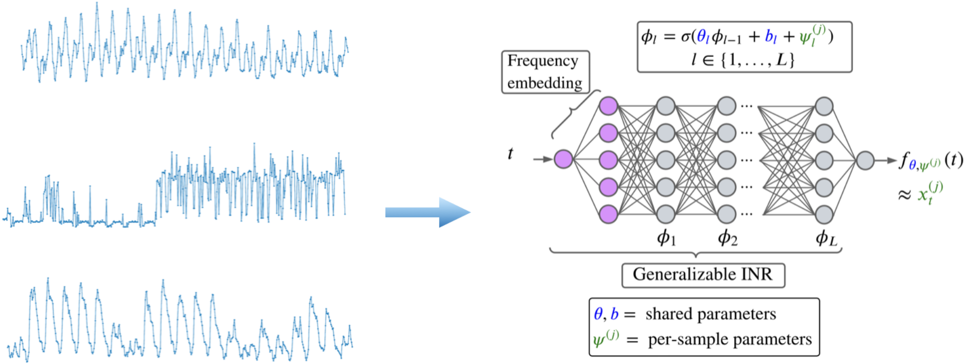
MoTM:
Mixture of TimeFlow Models

Vanilla Implicit Neural Representations (INRs)



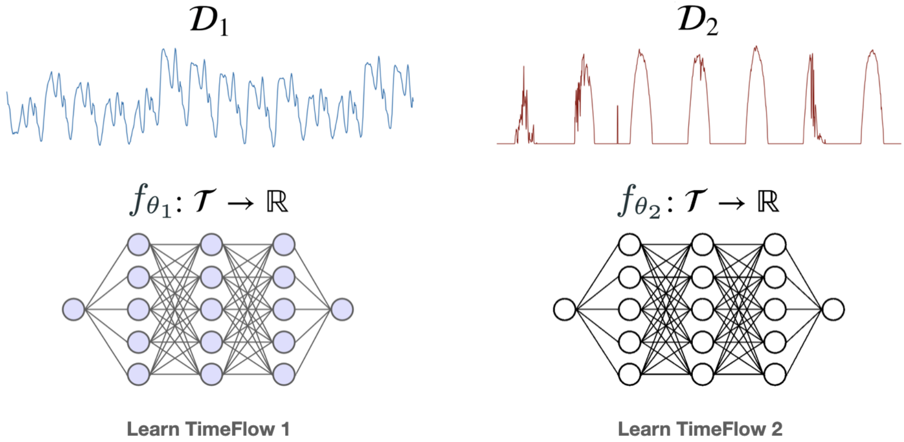
- Vanilla INRs are designed to fit a single instance

How to Deal with Datasets? → Generalizable INRs



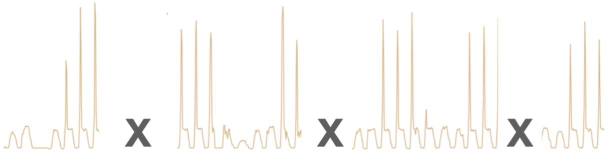
- TimeFlow [7] is a generalizable implicit neural representation that rapidly adapts to new time series

MoTM Training: Pretraining a Basis of TimeFlow Models on Each Dataset in $\mathcal{D}_{\text{train}}$



MoTM Inference: (i) Adapt, (ii) Orchestrate and (iii) Predict

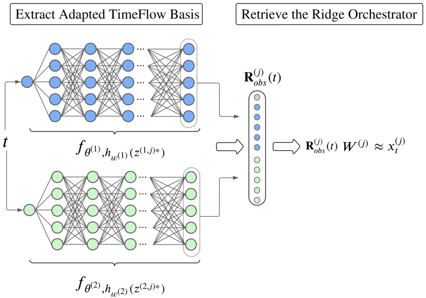
Let us consider a new (incomplete) time series $\mathbf{x_{new}}$



(i) **Adapt TimeFlows in the basis** : retrieve $f_{\theta_1, \Psi_1^{new}}$, $f_{\theta_2, \Psi_2^{new}}$ based on the observed context

(ii) **Orchestrate the TimeFlow's together**: fit a ridge regressor on top of the extracted representations

(iii) **Predict** for unobserved timestamps



Experiments

Training and In Domain Results

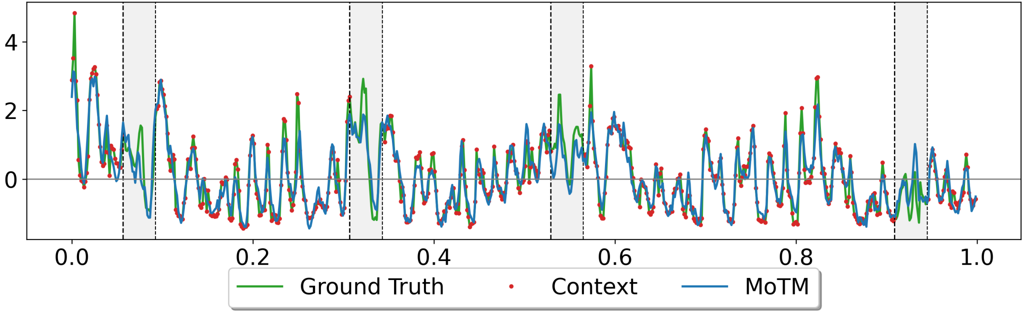
- MoTM was pretrained on three datasets: *Electricity*, *Solar*, and *Spanish W-T*.
- Evaluation of MoTM and baseline models across four imputation scenarios.

		Zero-shot		*Supervised*			*Statistical*	
		MoTM	**MOMENT**	**TimeFlow**	**BRITS**	**SAITS**	**Linear**	**Repeat**
MoTM In-Domain								
Electricity	*Point 1*	**0.196**	0.861	0.274	0.324	0.211	0.306	0.334
	Point 2	**0.229**	0.863	0.322	0.465	0.258	0.435	0.357
	Block 1	**0.257**	0.478	0.291	0.522	0.296	1.025	0.312
	Block 2	**0.259**	0.538	0.298	0.465	0.292	1.027	0.313
Solar	*Point 1*	0.083	0.857	0.085	0.072	0.077	**0.036**	0.265
	Point 2	0.092	0.858	0.097	0.083	0.130	**0.055**	0.281
	Block 1	0.253	0.755	0.257	0.308	0.361	0.883	**0.244**
	Block 2	0.256	0.781	0.258	0.314	0.355	0.889	**0.244**
Spanish W-T	*Point 1*	0.214	0.835	0.283	0.373	0.205	**0.169**	0.520
	Point 2	**0.253**	0.838	0.309	0.473	0.295	0.277	0.585
	Block 1	0.402	0.511	**0.391**	0.685	0.444	0.889	0.484
	Block 2	0.404	0.548	**0.396**	0.689	0.451	0.898	0.470
MoTM improvement		**0.0%**	**62.8%**	**10.8%**	**30.6%**	**12.6%**	**22.5%**	**32.2%**

Out of Domain Results

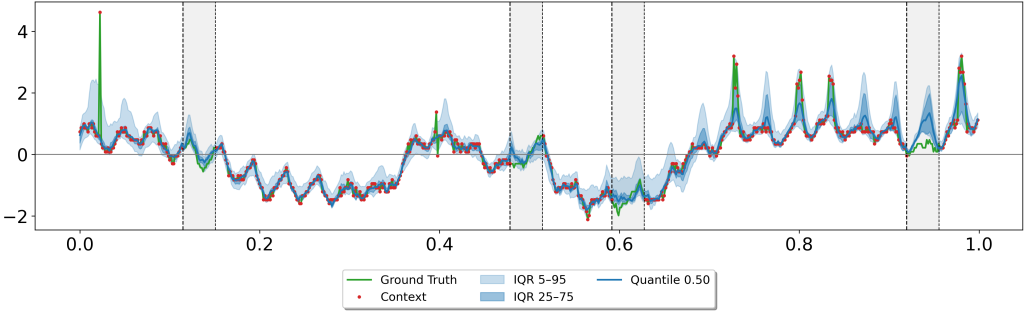
		Zero-shot		*Supervised*			*Statistical*	
		MoTM	**MOMENT**	**TimeFlow**	**BRITS**	**SAITS**	**Linear**	**Repeat**
MoTM Out-of-Domain								
Traffic	*Point 1*	0.246	0.770	0.240	0.267	**0.201**	0.287	0.379
	Point 2	0.294	0.774	0.291	0.374	**0.241**	0.421	0.416
	Block 1	0.313	0.478	0.389	0.415	**0.227**	0.983	0.340
	Block 2	0.318	0.521	0.395	0.431	**0.231**	0.985	0.341
ETTh1	*Point 1*	0.340	0.812	0.410	0.539	0.347	**0.334**	0.594
	Point 2	**0.389**	0.814	0.482	0.633	0.421	0.426	0.635
	Block 1	**0.490**	0.633	0.553	0.723	0.535	0.845	0.558
	Block 2	**0.488**	0.664	0.557	0.730	0.536	0.834	0.559
ETTh2	*Point 1*	0.442	0.806	0.489	0.533	0.422	**0.406**	0.763
	Point 2	0.496	0.806	0.521	0.610	0.486	**0.471**	0.805
	Block 1	0.609	0.704	0.631	0.691	**0.602**	0.761	0.738
	Block 2	**0.600**	0.704	0.619	0.722	0.610	0.760	0.716
Weather	*Point 1*	0.326	0.816	0.330	0.389	0.287	**0.260**	0.739
	Point 2	0.375	0.819	0.383	0.484	0.351	**0.323**	0.803
	Block 1	**0.524**	0.640	0.627	0.689	0.584	0.621	0.677
	Block 2	**0.527**	0.669	0.633	0.691	0.582	0.620	0.682
Spanish E	*Point 1*	0.235	0.818	0.329	0.311	0.189	**0.164**	0.678
	Point 2	0.286	0.823	0.376	0.425	0.285	**0.253**	0.738
	Block 1	0.507	0.622	**0.503**	0.675	0.536	0.606	0.644
	Block 2	0.503	0.649	**0.499**	0.675	0.535	0.604	0.633
MoTM improvement		**0.0%**	**40.3%**	**10.2%**	**24.0%**	**-5.7%**	**13.2%**	**31.1%**

Visual results on OoD datasets



OoD SpanishE dataset. Pointwise MoTM imputation on four one-day missing blocks.

Visual results on OoD datasets



OoD `Hog` *dataset.* Quantiles MoTM imputations on four one-day missing blocks.

Conclusion

Conclusion

- Proposed a flexible zero-shot imputation model
- Mixture approach yields consistent performance gains
- Out-of-domain performance comparable to the best supervised model, with major savings in time and data requirements

- Open question: how to best construct the MoTM training basis
- Potential improvements: stronger orchestrator/regressor

Thank you for listening

Link to the paper

Link to TimeFlow's paper

References i

References

[1] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, et al. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

[2] Wei Cao, Dong Wang, Jian Li, Hao Zhou, Yitan Li, and Lei Li. BRITS: Bidirectional Recurrent Imputation for Time Series. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[3] Jinguo Cheng, Chunwei Yang, Wanlin Cai, Yuxuan Liang, Qingsong Wen, and Yuankai Wu. Nuwats: a foundation model mending every incomplete time series. *arXiv preprint arXiv:2405.15317*, 2024.

[4] Abhimanyu Das, Weihao Kong, Rajat Sen, and Yichen Zhou. A decoder-only foundation model for time-series forecasting. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 10148–10167. PMLR, 2024.

References ii

[5] Wenjie Du, David Côté, and Yan Liu. SAITS: Self-attention-based imputation for time series. *Expert Systems with Applications*, 219:119619, 2023.

[6] Mononito Goswami, Konrad Szafer, Arjun Choudhry, Yifu Cai, Shuo Li, and Artur Dubrawski. MOMENT: A family of open time-series foundation models. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 16115–16152. PMLR, 2024.

[7] Etienne Le Naour, Louis Serrano, Léon Migus, Yuan Yin, Ghislain Agoua, Nicolas Baskiotis, Patrick Gallinari, and Vincent Guigue. Time Series Continuous Modeling for Imputation and Forecasting with Implicit Neural Representations. *Transactions on Machine Learning Research*, 2024.

[8] Mengxuan Li, Ke Liu, Jialong Guo, Jiajun Bu, Hongwei Wang, and Haishuai Wang. Imputeinr: Time series imputation via implicit neural representations for disease diagnosis with missing data. *arXiv preprint arXiv:2505.10856*, 2025.

[9] Gerald Woo, Chenghao Liu, Akshat Kumar, Caiming Xiong, Silvio Savarese, and Doyen Sahoo. Unified Training of Universal Time Series Forecasting Transformers. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pages 53140–53164. PMLR, 2024.