

Position: Current Benchmarking Hinders Real Progress in Deep Learning for Time Series Forecasting

Valentina Moretti¹, Ivan Marisca¹, Cesare Alippi^{1,3}, Andrea Cini^{1,2}

¹IDSIA, USI-SUPSI ²EPFL ³Politecnico di Milano



The State of Deep Learning for Time Series Forecasting

The field is stuck in a loop:

- ▶ Novel architectures claim SOTA results
- ▶ Results are quickly dismissed by newer architectures
- ▶ Contradictory findings across studies
- ▶ Simple models often match complex ones



The problem goes beyond common pitfalls in empirical evaluation.

Position Statement

Our current approach to building and comparing deep learning architectures for time series forecasting **overlooks fundamental design dimensions** and their interplay. As a result, it **fails in explaining observed empirical results** and **supporting progress** in the field.

Are benchmarks measuring actual progress?

- ▶ Current benchmarks compare architectures that differ in multiple design choices → **cannot attribute performance gains** to specific components
- ▶ Observed improvements may stem from failure in factoring out differences in design, rather than from methodological advances

Four Key Design Dimensions

Benchmarks must disentangle the effects of these dimensions

D1. Model Configuration

Local vs. global vs. hybrid approaches

- ▶ Shared vs. series-specific parameters
- ▶ Inductive vs. transductive settings

D2. Preprocessing & Exogenous Variables

Data transforms and additional inputs

- ▶ Scaling, detrending, seasonality
- ▶ Calendar features, covariates

D3. Temporal Processing

Operators to model temporal dependencies

- ▶ Attention, RNN, TCN, MLP, ...
- ▶ Main focus of recent research

D4. Spatial Processing

Operators for inter-series dependencies

- ▶ Spatial attention, graph-operators

Experimental Evidence

Problem: In current benchmarks, compared architectures differ in **multiple design choices at once** making it impossible to attribute performance to any single component.

Our approach: We show the **impact of overlooked design dimensions** on SOTA models by varying one choice at a time.

- D1** Benchmarks mix hybrid and global models. We run each in **both configurations** → the choice alone **can flip rankings**.
- D2** Some models use exogenous variables, others don't. We add/remove them → **large performance shifts**.
- D3** We properly configure simple baselines (MLP, RNN) → they **match or beat** complex architectures. Not factoring out design dimensions **misleads conclusions** drawn from comparisons.
- D4** We remove the “key” spatial component from SOTA models (e.g. spatial att. → MLP) → performance **stays the same or improves**.

Call to Action

1. Better Benchmarks

- ▶ Isolate effects of proposed designs
- ▶ Keep other dimensions fixed

2. Explicit Documentation

- ▶ Report all design choices
- ▶ **Forecasting Model Card**

Forecasting Model Card



Forecasting Model Card Builder

Forecasting model cards are structured documentation for time series forecasting models. We built this tool based on our publication [Position: Current Benchmarking Hinders Real Progress in Deep Learning for Time Series Forecasting \(ICML 2026\)](#). Credits: Inspired from [Hugging Face Model Cards Writing Tool](#).

[Download Markdown](#)[Copy Markdown](#)

Sections

Section 1 of 11

[Model Overview](#)[Model Setting](#)[Model Configuration](#)[Preprocessing and Exogenous Variables](#)[Temporal Processing](#)[Spatial Processing](#)[Risks and Limitations](#)[Training Details](#)[Environmental Impact](#)[Citation](#)[Authors and Contact](#)

SECTION 1 OF 11

Core metadata that sets the card header and frontmatter.

Model Overview

Model Name

Model Description

Provide a short summary of what the model does.

Developed By

Comma-separated if multiple.

License

Library

Task

LIVE OUTPUT

Rendered Preview

Model Name

Table of Contents

- [Model Overview](#)
- [Model Setting](#)
- [Model Configuration](#)
- [Preprocessing and Exogenous Variables](#)
- [Temporal Processing](#)
- [Spatial Processing](#)
- [Risks and Limitations](#)
- [Training Details](#)
- [Environmental Impact](#)
- [Citation](#)
- [Authors and Contact](#)

Model Overview

Model Setting

Model Configuration

Preprocessing and

Thank you!

Questions?



`valentina.moretti@usi.ch`