Foundational Time Series Models:

the future of forecasting, or just hype?

ARIMA → ??? → TimeGPT, TimesFM, Moirai, Chronos

TimeGPT

TimesFM

Moirai

salesforce

Chronos

amazon

PyData Global, 2024
Ahad Shoaib, Salesforce

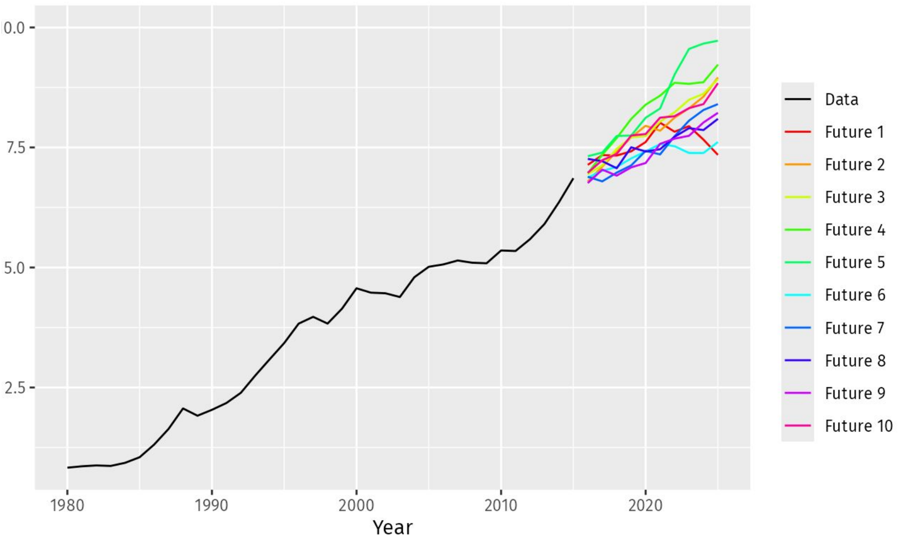
Agenda

- ❖ Time Series Forecasting 101
- ❖ The journey from ARIMA → Foundational models
- ❖ How should you use foundational models?
  - ➢ *and how you **shouldn't***
- ❖ What's next?

Time Series Forecasting 101

Problem: Estimating future values using observations

generally follows the standard ML lifecycle: data→ fit → predict → evaluate → interpret



A few use cases...

Sales Forecasting

- Retail Store Stocks (ex. Walmart)
- Finance

Infrastructure Management ★

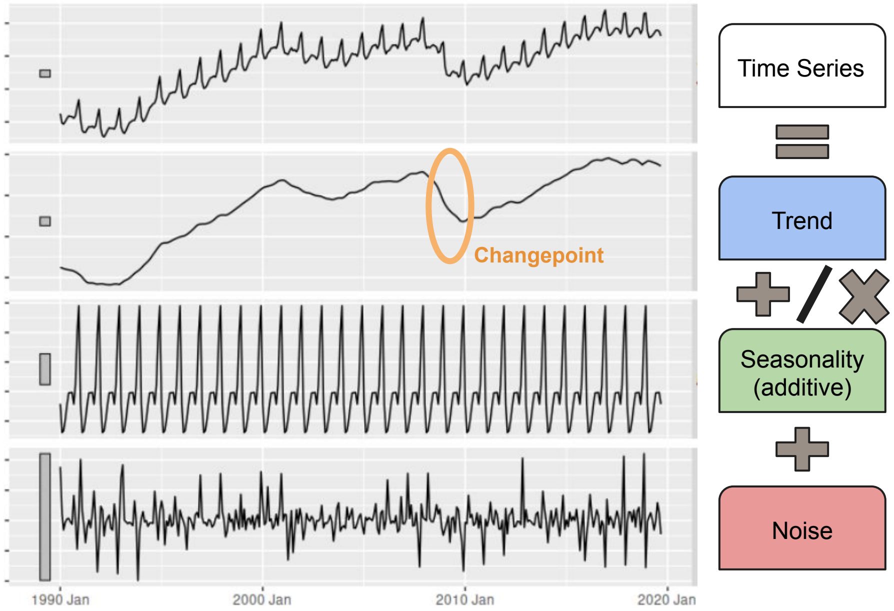
- Capacity Planning
- Cost Optimization

Customer Support

- Ticket Volume Prediction
- Support Call Duration Prediction

★ *Infrastructure Data Science team @ Salesforce*

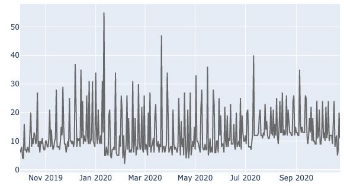
Context: Understanding the Language of Time Series

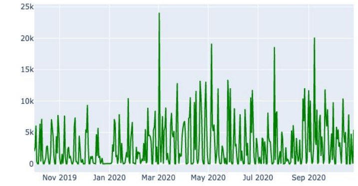


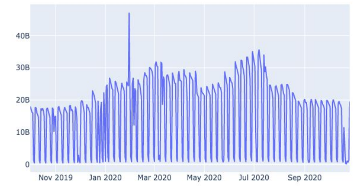
<u>What does forecasting generally look like?</u>

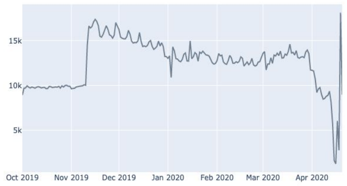
Forecast(t+1) = Trend(t) + Seasonality(t) + Error(t)

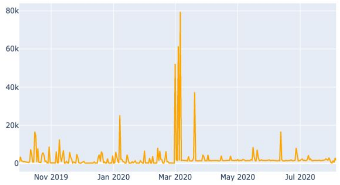
Challenge: lots of different shapes, sizes, granularities

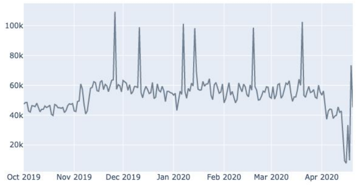


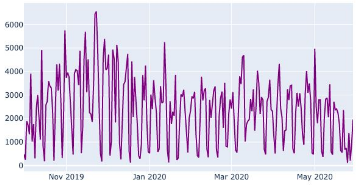


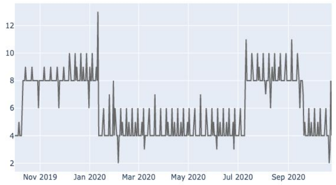


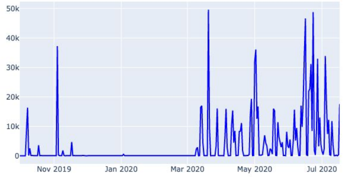


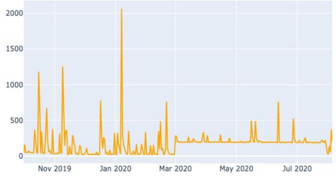


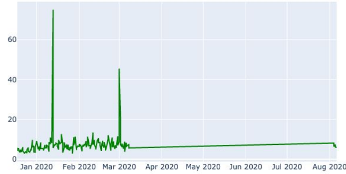












Solution: The Forecasting Community is a bit… divided

Are Transformers Effective for Time Series Forecasting?

Ailing Zeng[1]*, Muxi Chen[1]*, Lei Zhang[2], Qiang Xu[1]
[1]The Chinese University of Hong Kong
[2]International Digital Economy Academy (IDEA)

Yes, Transformers are Effective for Time Series Forecasting (+ Autoformer)

Published June 16, 2023

Update on GitHub

elisim Eli Simhayev guest
kashif Kashif Rasul
nielsr Niels Rogge

★ [Do We Really Need Deep Learning Models for Time Series Forecasting?] [Oct. 2021]

★ [Are Transformers Effective for Time Series Forecasting?] [Aug. 2022]

★ [Yes, Transformers are Effective for Time Series Forecasting (+ Autoformer)] [June 2023]

★ [Are Language Models Actually Useful for Time Series Forecasting?] [June 2024]

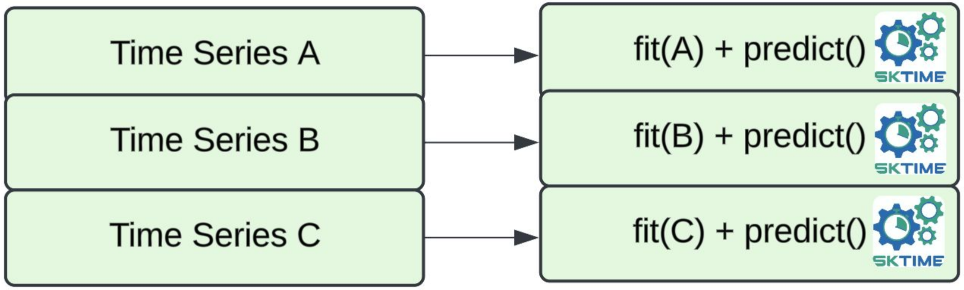
★ **Foundational Time Series Models: The Future, or Hype?** **[This Talk!]**

salesforce

A journey from ARIMA to foundational models

Overview: Local vs Global Forecasting Models

Local Models

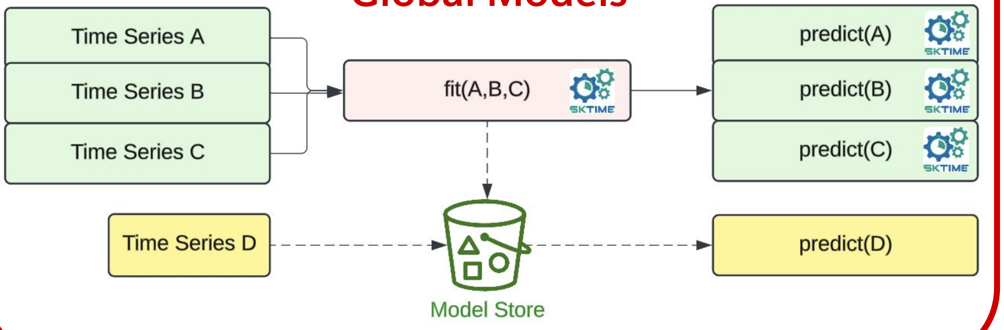


ARIMA, ETS, Prophet, ...

- Simple to build & interpret
- Easy to tweak individually with prior knowledge

But.. not always accurate @ scale

Global Models



LightGBM, DeepAR, ...

- More expressive
- Accurate @ scale
  - 🏆 M5 Competition

But... require feature engineering + some ML infrastructure

Blueprint for Global Time Series Model Architecture

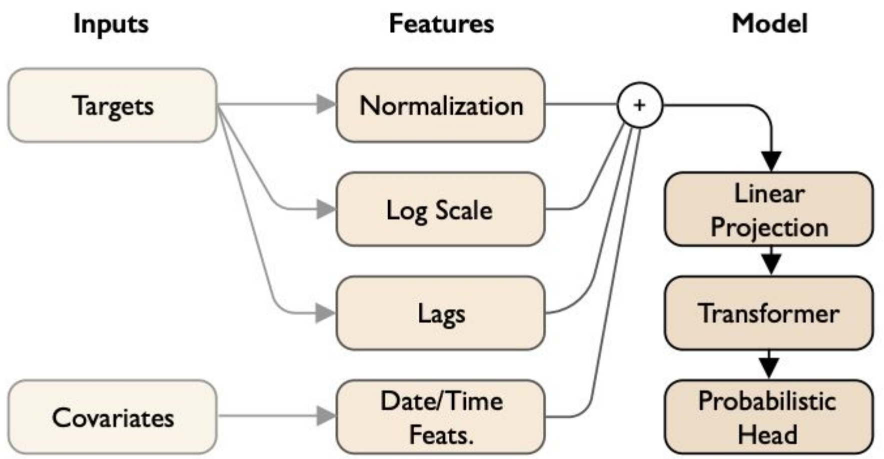


Figure 3: Data flow of a standard pipeline for probabilistic time series forecasting (Salinas et al., 2020).

Pushing the Limits of Pre-training for Time Series Forecasting in the Cloudops Domain (Salesforce Research, 2023)

<2017: Forecasting is slowly maturing, NLP enters new era

***Forecasting:** classical methods prevail*

Latest M3 Competition:
Oct. 2000

LightGBM

forecast

PROPHET
Forecasting at scale.

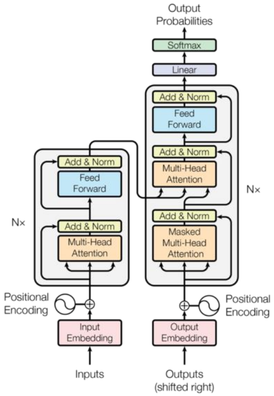
DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks

David Salinas, Valentin Flunkert, Jan Gasthaus
Amazon Research
Germany
<dsalina,flunkert,gasthaus@amazon.com>

amazon

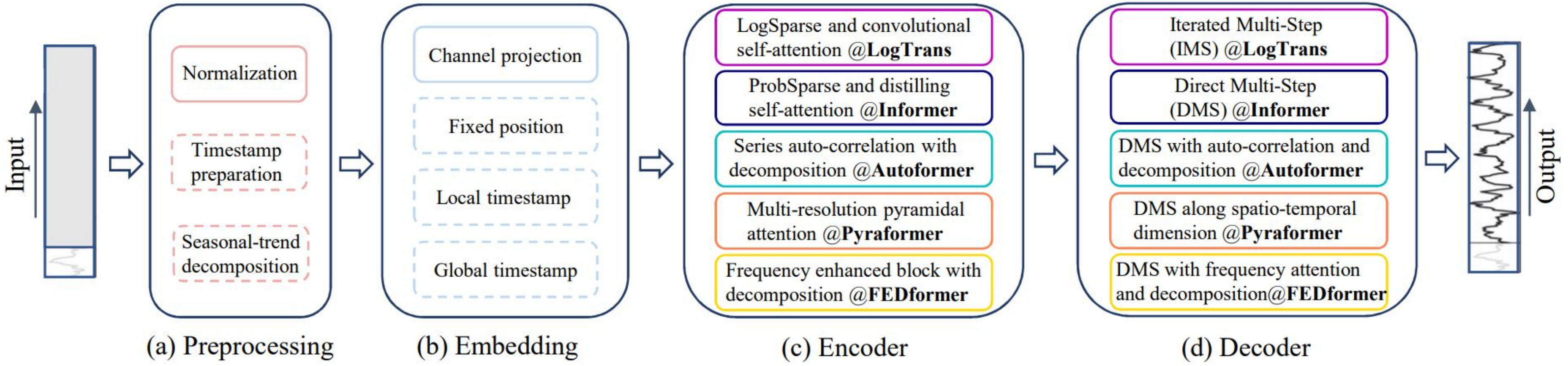
June 2017: Transformers released!

Attention Is All You Need



2017-2022: Python Adoption & Transformers

Python + Time Series: sktime, statsforecast, Darts, kats, orbit, greykite, ...



DLinear/NLinear: But, a simple MLP + MovingAverage + “RevIN” beats transformers...

- *ARIMA’s still in the game!*

image from Are Transformers Effective for Time Series Forecasting?

A Time Series is Worth 64 Words: Long-term Forecasting with Transformers

Yuqi Nie[1]*, Nam H. Nguyen[2]*, Phanwadee Sinthong[2], Jayant Kalagnanam[2]
[1]Princeton University [2]IBM Research
ynie@princeton.edu, nnguyen@us.ibm.com, Gift.Sinthong@ibm.com, jayant@us.ibm.com

2023: PatchTST released

Intuition:

- [2020 LogSparse]: it's OK to ignore timesteps close to each other, as long as you still include a long enough context window
- **NLP:** BERT tokenizes sub-words (groups of characters) instead of characters
- **Vision:** "An Image is Worth 16x16 words"

Question: "is there a way to avoid throwing values while maintaining a long look-back window"?

- **Answer:** Patches!

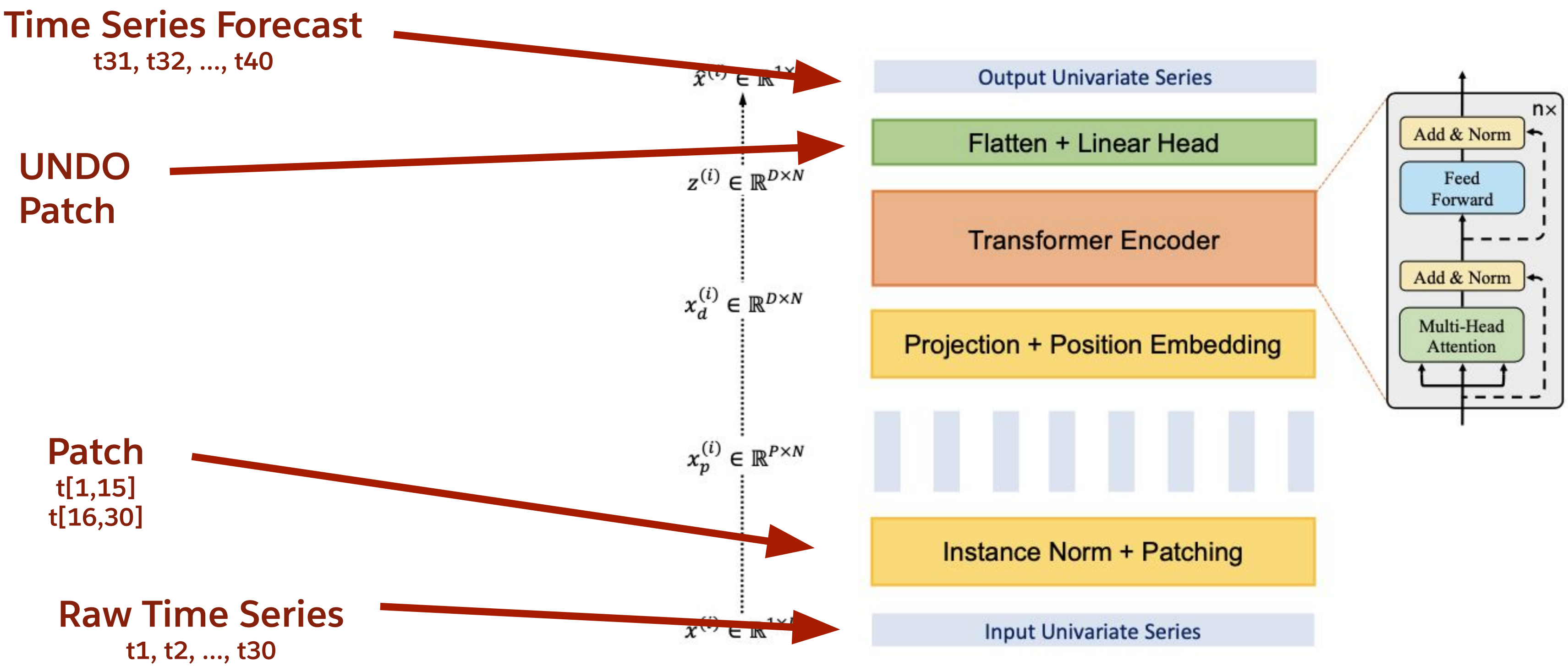
Result: State of the Art →

T	model	MAPE	CRPS	Rank
	PatchTST	0.860	0.563	5.897
	Moirai_large	0.856	0.585	6.072
	Moirai_base	1.101	0.604	6.155
	TFT	1.116	0.596	7.031
	Moirai_small	0.909	0.636	8.423
	Chronos_large	0.930	0.629	8.814

Foundational Building Block: Patching

A TIME SERIES IS WORTH 64 WORDS: LONG-TERM FORECASTING WITH TRANSFORMERS

Yuqi Nie[1]*, **Nam H. Nguyen**[2]*, **Phanwadee Sinthong**[2], **Jayant Kalagnanam**[2]
[1]Princeton University [2]IBM Research
ynie@princeton.edu, nnguyen@us.ibm.com, Gift.Sinthong@ibm.com, jayant@us.ibm.com



<u>2024:</u> Transformer-Based Foundational Models (ex. Moirai)

Learnings from NLP*: Less inductive bias/structure + more data/compute

Architecture: [1] Transformer + [2] Patching + [3] Multi-Granularity + [4] Diverse Datasets

UNDO Patch

Multi-Patching

*great talk on the intuition behind the evolution of transformers by [Hyung Won (Stanford CS25)]

The Recipe for Foundational Time Series Models in 2024

Big Data + Transformers OR LLMs

Big Data

- Wikipedia + Synthetic Data
- LOTSA: Curated Datasets
- Time Series Pile

Transformers
+ patching

- TimeGPT*
  *closed source
- TimesFM
- Moirai
- MOMENT
- Tiny Time Mixers*
  *based on TSMixer backbone

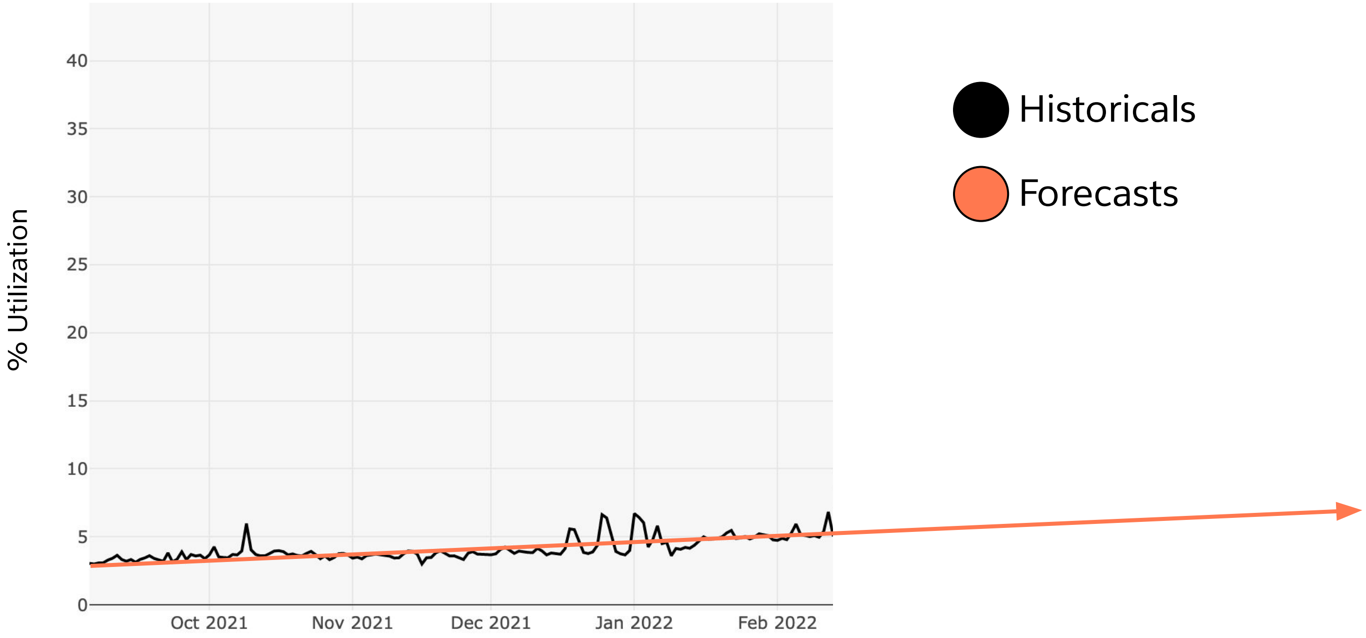
LLMs

- Lag-Llama
- Chronos
- GPT4TS

salesforce

So, how should you use foundational time series models?

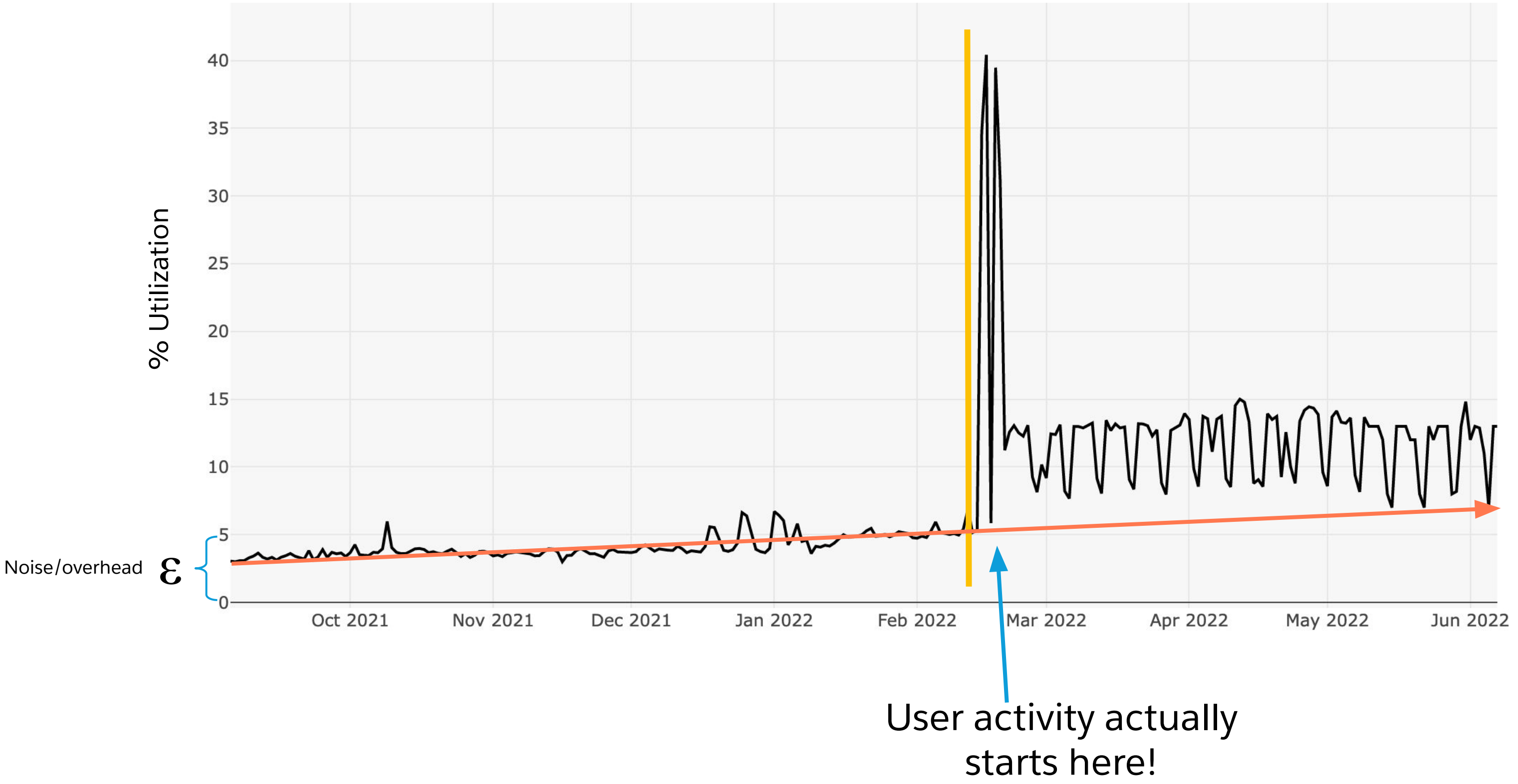
Wow! Can foundational models really predict the future?



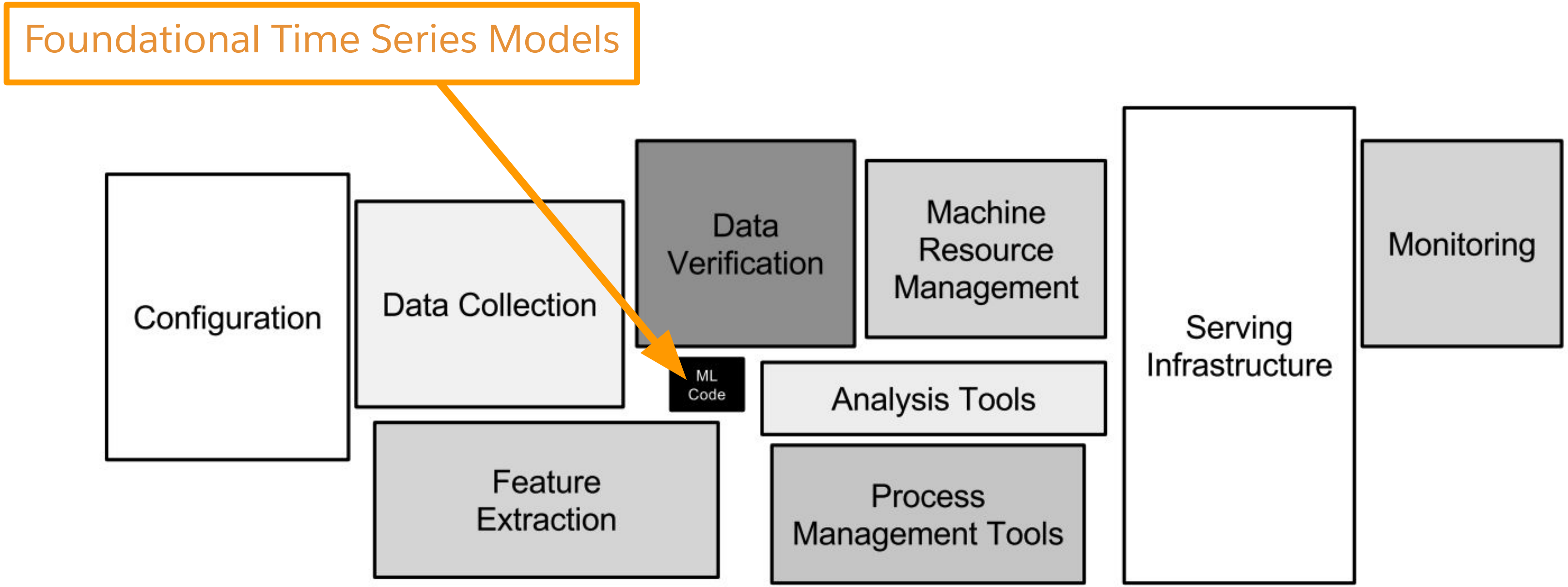
*there is usually a 0-5%~ constant utilization overhead

No, you still need data + some external knowledge

Solution ideas: (1) use "similar" time series, (2) dynamic hierarchies, (3) human-based [drivers]()



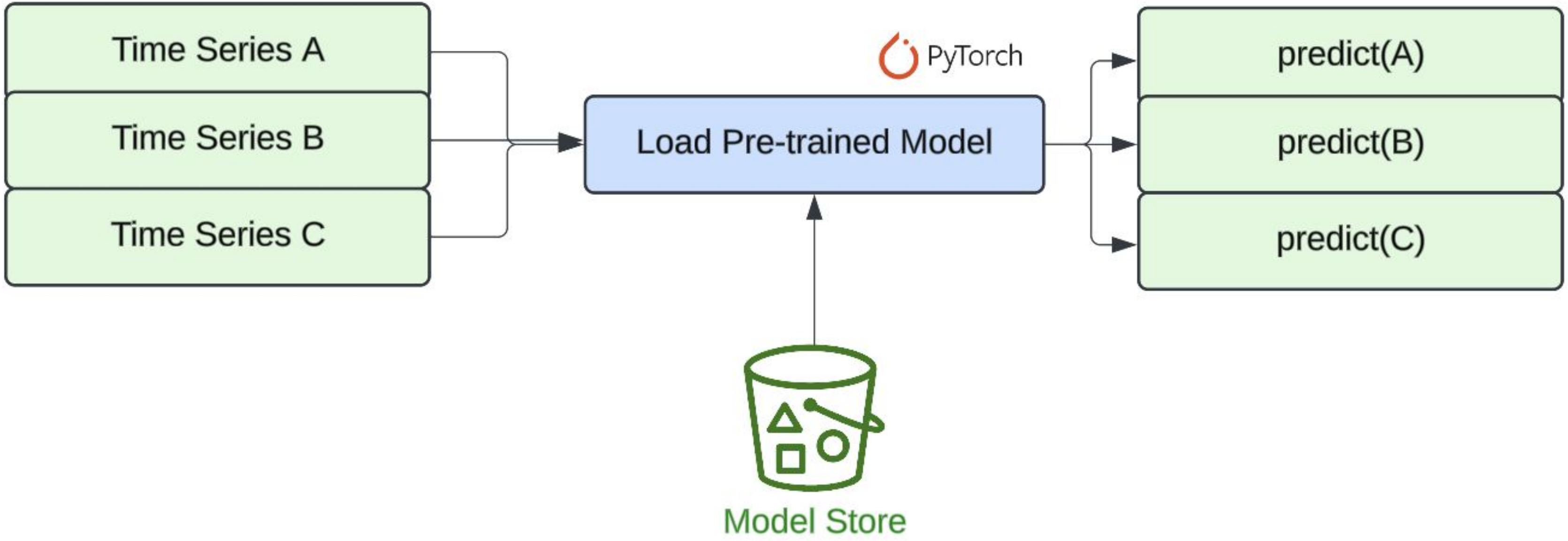
Ok, but can I just replace my Forecasting team & ML Platform with these models? Not yet!



Interested in ML platforms? Here's a blog on how [we built our Forecasting Platform]()

What does this mean for me, as a forecaster?

- It's still a ML model, but without training – data in, forecasts out!
- **TLDR:** Higher accuracy out of the box, less training/tuning effort, better "cold start"



When should I train my own foundational model?

Poor backtesting results? Have lots of time series? Tired of feature engineering?

Forecasting cloud infrastructure?
Consider adding more specialized data!

(Moirai) LOTSA Training Dataset Distribution

Table 2. Key statistics of LOTSA by domain.

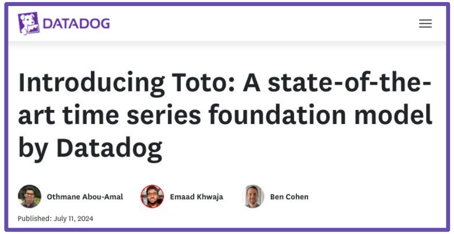
	Energy	Transport	Climate	CloudOps	Web	Sales	Nature	Econ/Fin	Healthcare
# Datasets	30	23	6	3	3	6	5	23	6
# Obs.	16,358,600,896	4,900,453,419	4,188,011,890	1,518,268,292	428,082,373	197,984,339	28,547,647	24,919,596	1,594,281
%	59.17%	17.73%	15.15%	5.49%	1.55%	0.72%	0.09%	0.10%	0.01%

Table 3. Key statistics of LOTSA by frequency.

	Yearly	Quarterly	Monthly	Weekly	Daily	(Multi) Hourly	(Multi) Minute-level	(Multi) Second-level
# Datasets	4	5	10	7	21	31	25	2
# Obs.	873,297	2,312,027	11,040,648	18,481,871	709,017,118	19,875,993,973	7,013,949,430	14,794,369
%	0.003%	0.008%	0.040%	0.067%	2.565%	71.893%	25.370%	0.054%

Case Study: Toto = LOTSA (27B) + CloudOps Data (750B)

- Fine-tuning/re-training with a larger set of domain specific dataset
  - *Better than (1) Zero-shot or (2) Domain-specific model*


DATADOG

Introducing Toto: A state-of-the-art time series foundation model by Datadog

Othmane Abou-Amal Emaad Khwaja Ben Cohen

Published: July 11, 2024

	Zero Shot				Full Shot		
	Toto	$\text{Moirai}_{\text{Small}}$	$\text{Moirai}_{\text{Base}}$	$\text{Moirai}_{\text{Large}}$	iTransformer	TimesNet	PatchTST
MAE	**0.312**	0.357	*0.341*	0.350	0.357	0.378	0.362
MSE	**0.265**	0.328	*0.315*	0.330	0.328	0.336	0.333

LOTSA + DD Dataset + Transformer

LOTSA

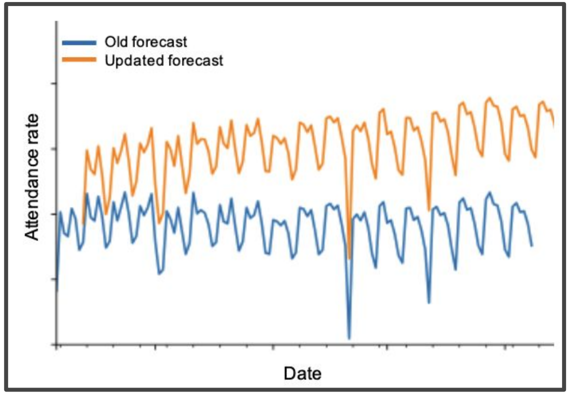
DD Dataset

- *We've found similar improvements from our own infrastructure usage data*

Model stability: critical for automated forecasting

- **Training instability** observed in large global models – critical for automated re-training
  - Generally not as important for “pre-trained” models, but...

Resilient Neural Forecasting Systems



- **Inference stability**
  - Anomaly Detection as preprocessing → greatly improved uncertainty intervals
  - “# samples” generated in Forecasting == “Temperature” in LLMs

What's next?

Continuing Research @ Time Series Forecasting

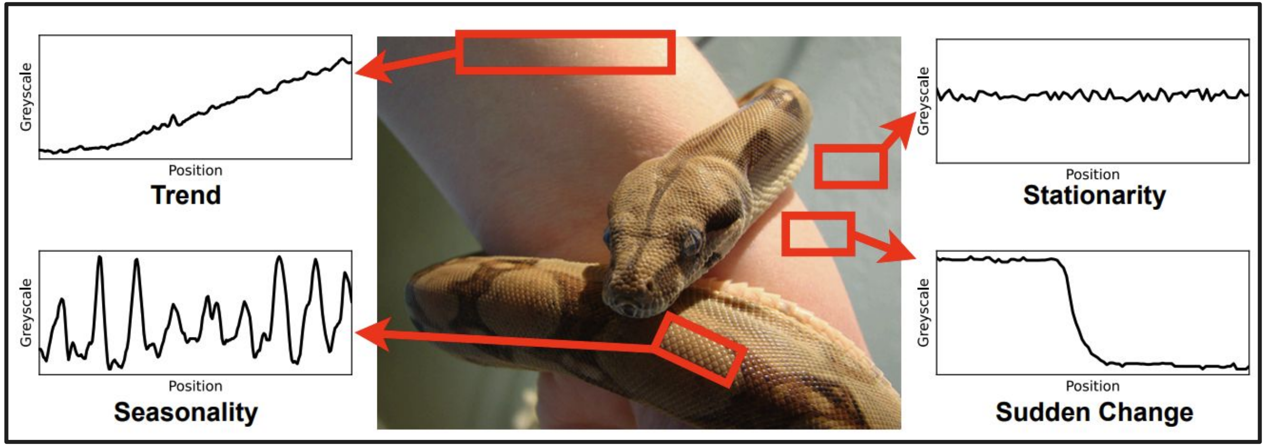
Mostly advancements from October 2024+

1. More open source datasets

- Time Series Pile
- GIFT-Eval
- Time-300B

2. Better models

- MoE-based (Time-MoE, Moirai-MoE)
- Mamba-based (CMMamba, Foundational)
- Vision-based (VisionTS)

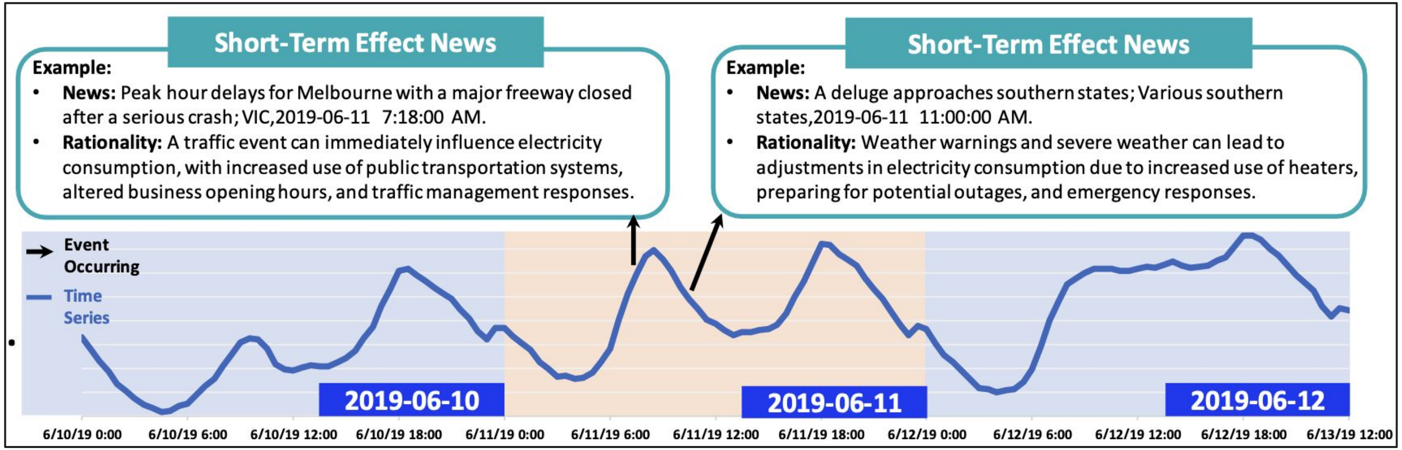


VisionTS: Visual Masked Autoencoders Are Free-Lunch Zero-Shot Time Series Forecasters

Multimodal Explainability: Time Series + Language

3. Time Series Understanding

- **Applications:** hierarchical time series correlations, root cause analysis (changepoints)
- LLMs + News Data [1][2]
- General Reasoning [3][4]



(Xinlei Wang et al.) From News to Forecast: Integrating Event Analysis in LLM-Based Time Series Forecasting with Reflection

Foundational Time Series Models

the future?

hype?

Foundational Time Series Models

the future!

hype!

Thank you

Interested in chatting about time series? Reach out on Linkedin!

Useful Links

- **Models:**
  - Moirai + Moirai-MoE: https://github.com/SalesforceAIResearch/uni2ts
  - Time-MoE: https://github.com/Time-MoE/Time-MoE
  - TimesFM: https://github.com/google-research/timesfm
  - Chronos: https://github.com/amazon-science/chronos-forecasting
  - Moment: https://github.com/moment-timeseries-foundation-model/moment
  - TimeGPT API: https://docs.nixtla.io/docs/getting-started-timegpt_quickstart
- **Datasets**
  - LOTSA: https://huggingface.co/datasets/Salesforce/lotsa_data
  - Time Series Pile: https://huggingface.co/datasets/AutonLab/Timeseries-PILE
  - Time-300B: https://huggingface.co/datasets/Maple728/Time-300B
  - GIFT-Eval: https://huggingface.co/datasets/Salesforce/GiftEval
- **Model Leaderboards:**
  - https://huggingface.co/spaces/Salesforce/GIFT-Eval
  - https://github.com/thuml/Time-Series-Library