When Zero-shot Beats Best-in-class

A DEEP DIVE INTO TIME SERIES FOUNDATION MODELS

Mihail “Mike” Mihaylov

May 2026

innovation forward

sirris

The innovation companion of the technology industry

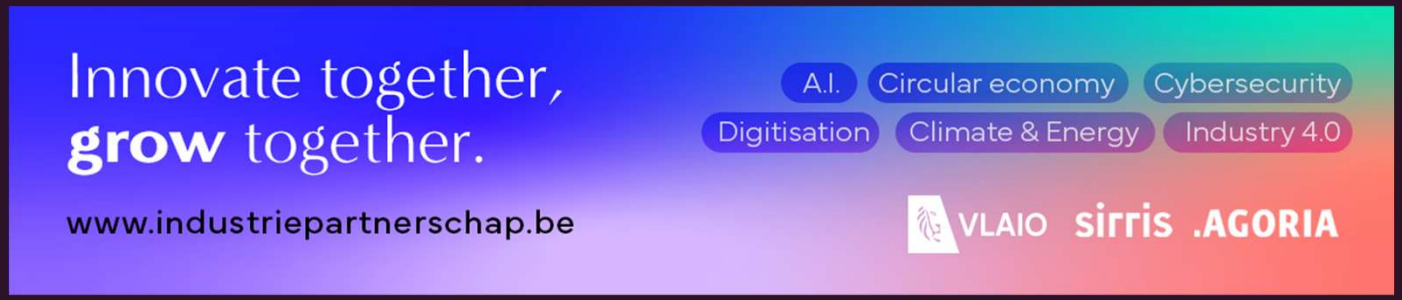
Sirris is a joint industry innovation center

Boosting the technological competitiveness
of the Belgian industrial sector.

Acknowledgements

The material for this seminar was developed in collaboration with the FAIR program's special project "Trustworthiness of Generative AI in Industrial R&D".

And with the support of Agentschap Innoveren & Ondernemen.



Mihail “Mike” Mihaylov

SENIOR GENAI ENGINEER | AI PRODUCT OWNER | ENERGY EXPERT

https://www.linkedin.com/in/mike-mihaylov/

mihail.mihaylov@sirris.be

18 years in academia – applied AI research

10 years in industry – Smart Energy SaaS

Solo entrepreneur as side profession

Expert evaluator for EC funding programs

Fascinated by the GenAI tech

Currently working with timeseries foundation models (TFMs) and GenAI trustworthiness

Core idea

GENAI FORECASTING MODELS ARE DISRUPTING THE FORECASTING DOMAIN

Current situation:

- Companies record timeseries data from machines, processes,… Most data sits unused.
- Forecasting is useful, but expensive: need ML engineers & SW developers.

Timeseries Foundation models* change the equation:

- Free models that forecast out-of-the-box. No training or tuning.
- Comparable accuracy to ML models
- Lower the cost barrier of forecasting use cases

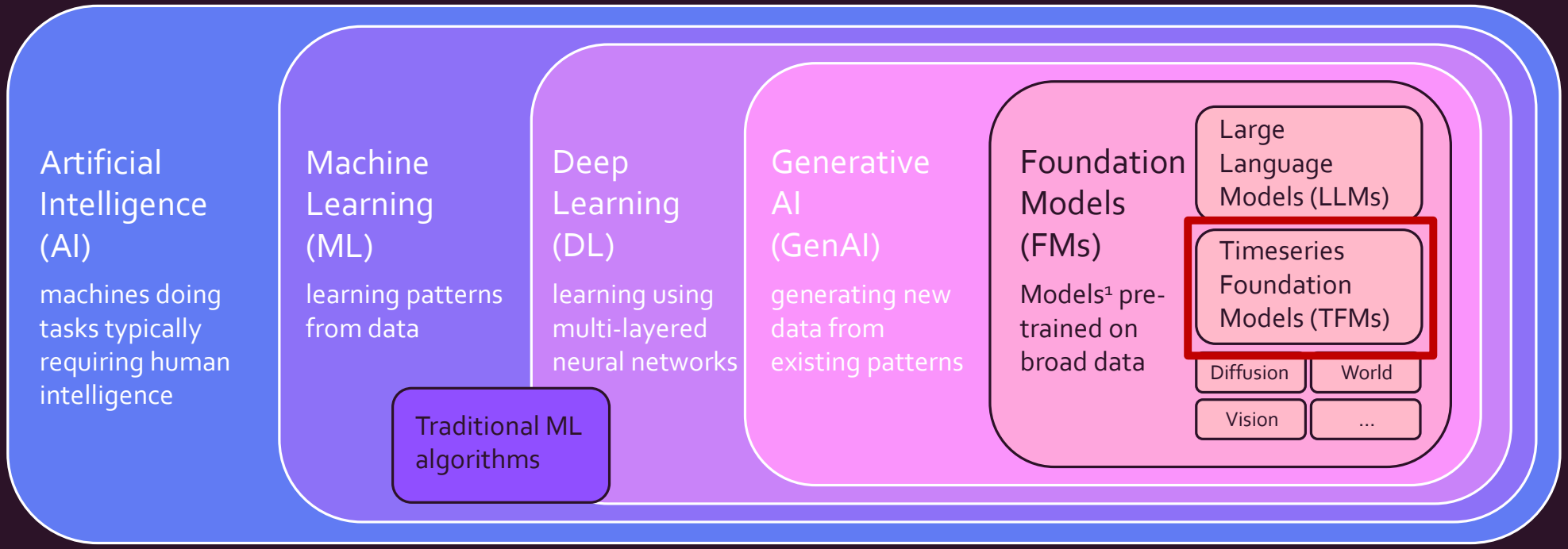
* Transformer-based models pre-trained on broad time-series data.

Timeseries Forecasting

TRADITIONAL ML VS FOUNDATION MODELS

innovation
forward

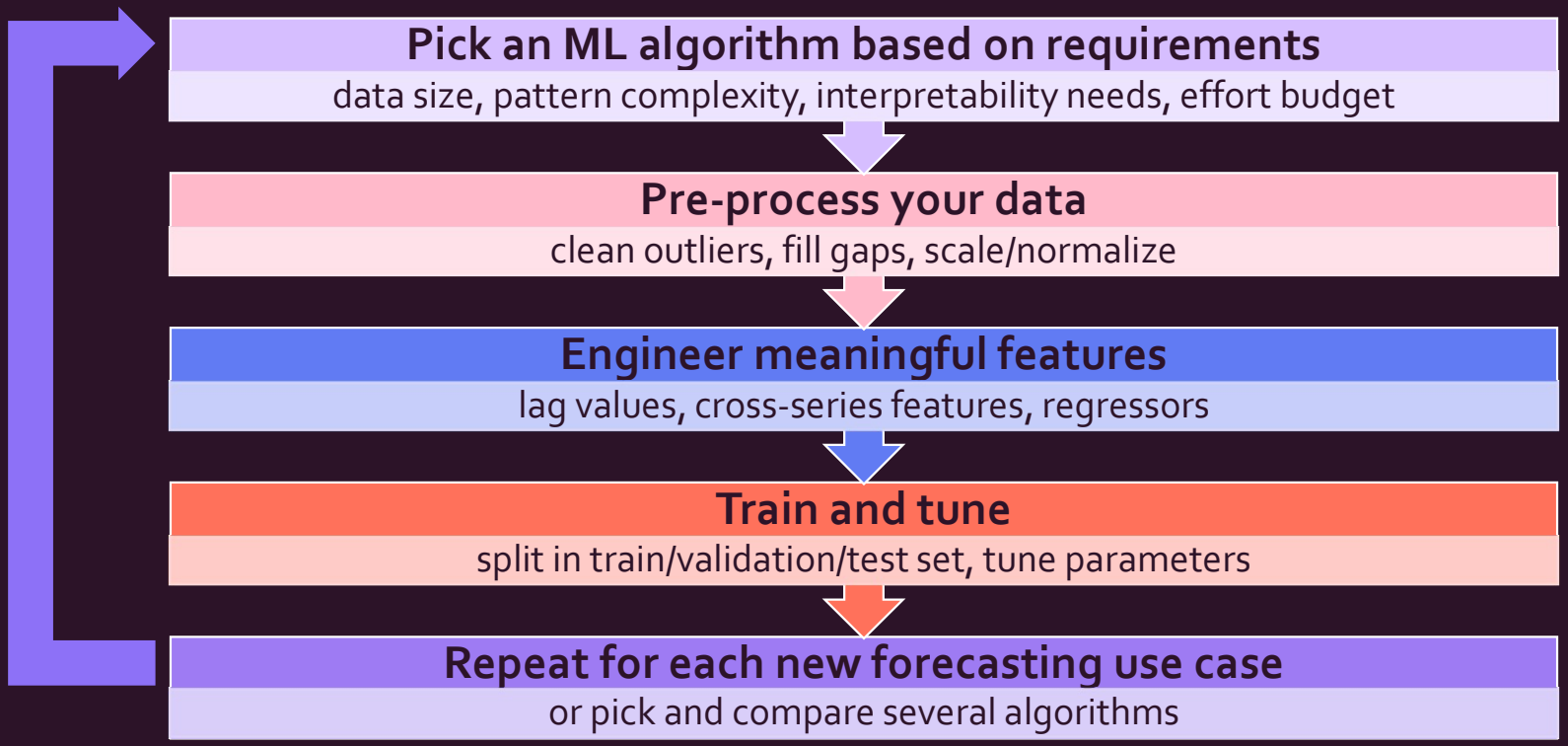
Hierarchical Taxonomy of AI Disciplines and Concepts



[1] model = neural network architecture + trained weights

Traditional Machine Learning Approach

FOR TIMESERIES FORECASTING



Can be:

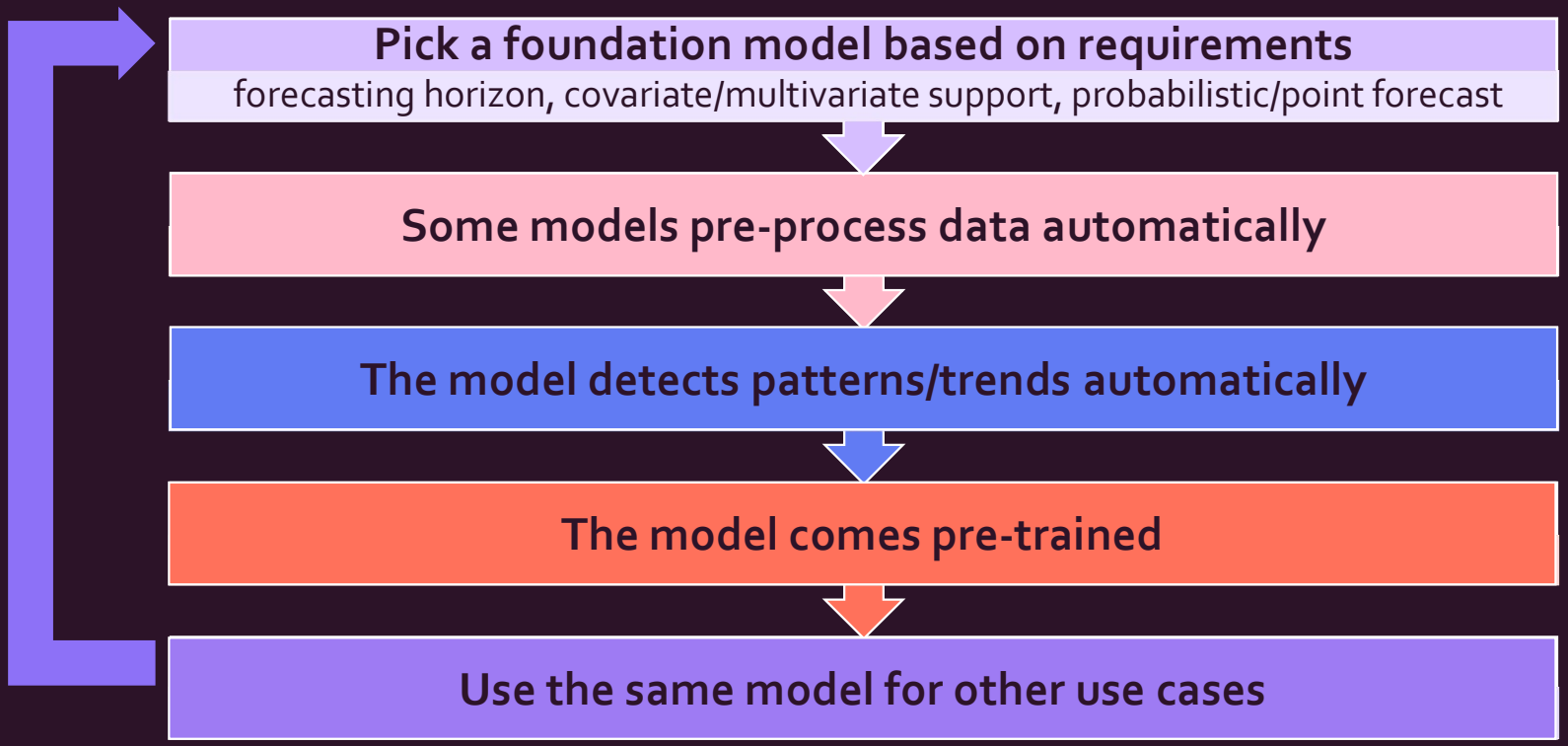
- Accurate
- Deterministic
- Interpretable

Requires:

- ML expertise
- Time
- Per-case effort

Novel Foundation Model Approach

FOR TIMESERIES FORECASTING



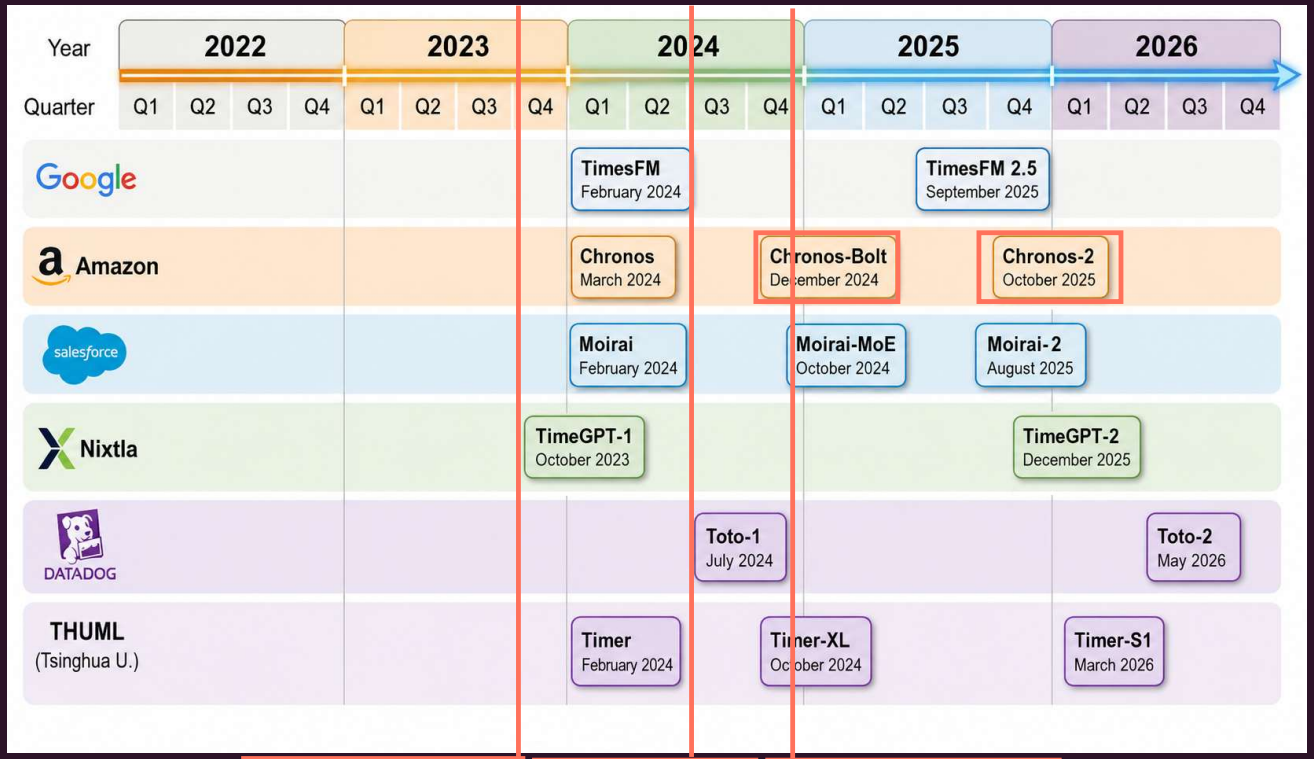
Pros:

- Accurate
- Versatile
- Low effort

Cons:

- Black box
- Expensive Fine-tuning

Timeline of Timeseries Foundation Models

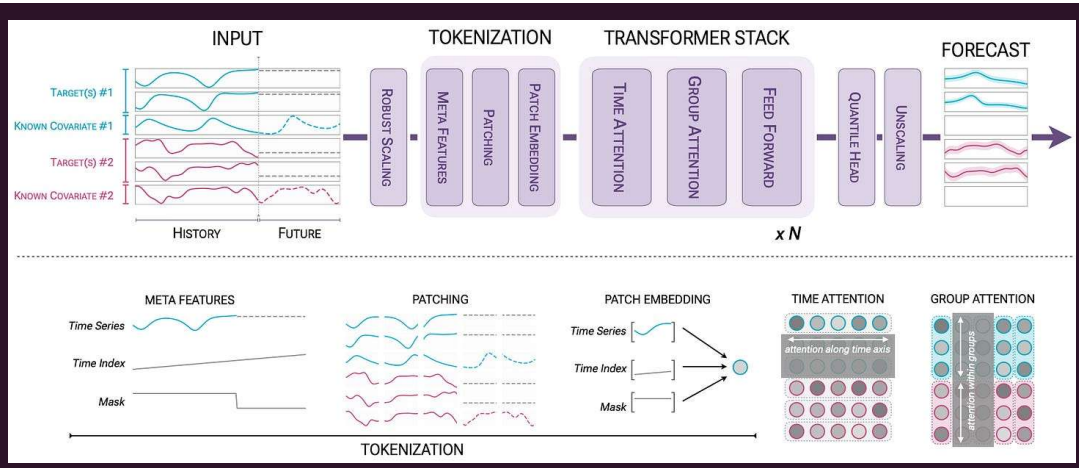


first academic publications

reasonable accuracy

comparable to classical ML

Chronos-2



Released October 2025 by Amazon Science

Supports uni/multi-variate, past/future covariates, percentiles (P1, P5, P10, …)

Topped forecasting leaderboards (Gift-Eval, Fev-Bench) up to April 2026

Input is normalized, transformed and masked -> **resilient to outliers and missing data**

Pre-trained on a combination of real-world and large-scale synthetic datasets

Apache 2 license. Suitable for commercial use.

Experimental results

WITH CHRONOS-2

innovation
forward

Use case 1: Day-ahead load forecasting

COMPARED TO ELIA'S OWN FORECASTER

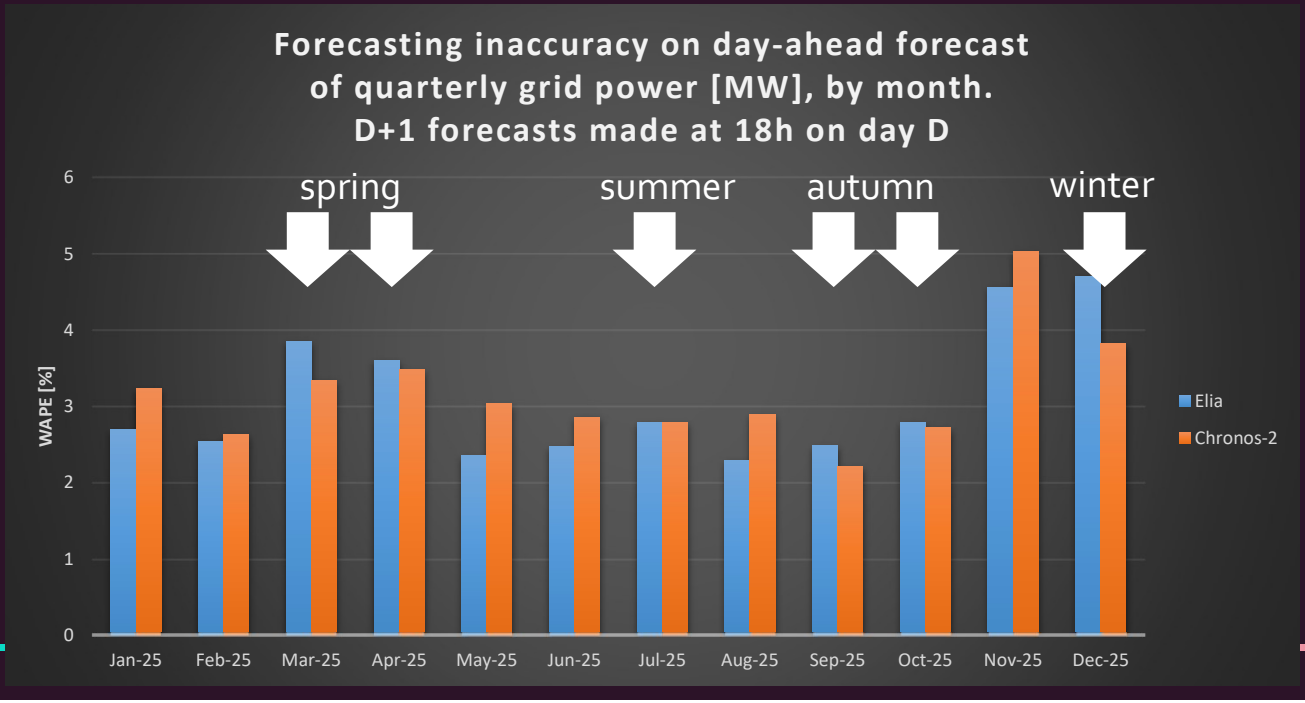
Error metric: WAPE

$$\mathrm{WAPE} = \frac{\sum_{t=1}^{N} |F_t - A_t|}{\sum_{t=1}^{N} |A_t|}$$

Weighted Absolute Percentage Error
"Lower percentage is better"

Accuracy of Chronos-2 ≥ Elia's forecaster on half of 2025, across seasons.

Only univariate forecasting with 30 days historical data. No features. No training.

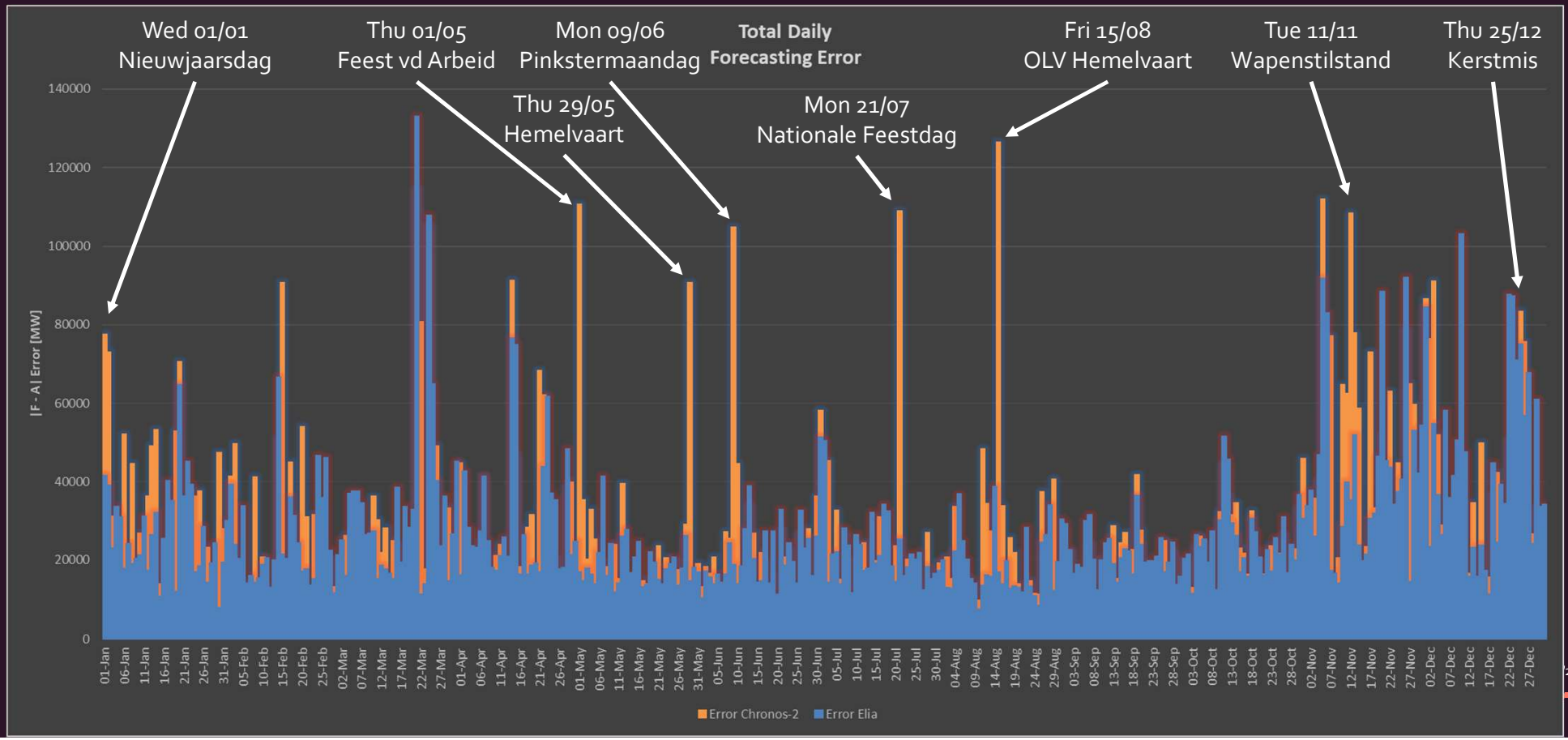


WAPE over 1y:
Elia = 3.12%
Chronos-2 = 3.19%

innovation forward

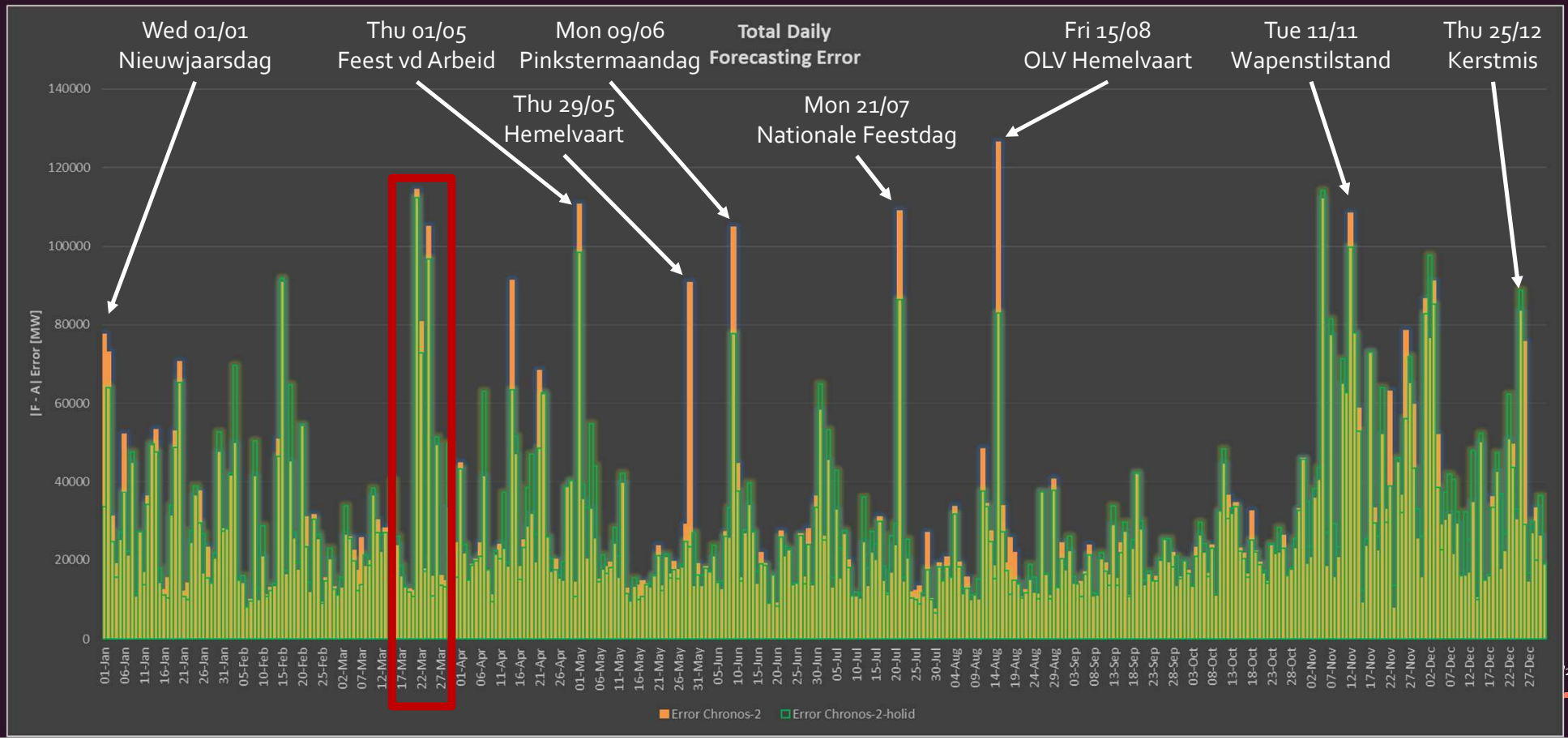
Some errors are predictable – on weekday holidays

BUT NOT ALL ERRORS

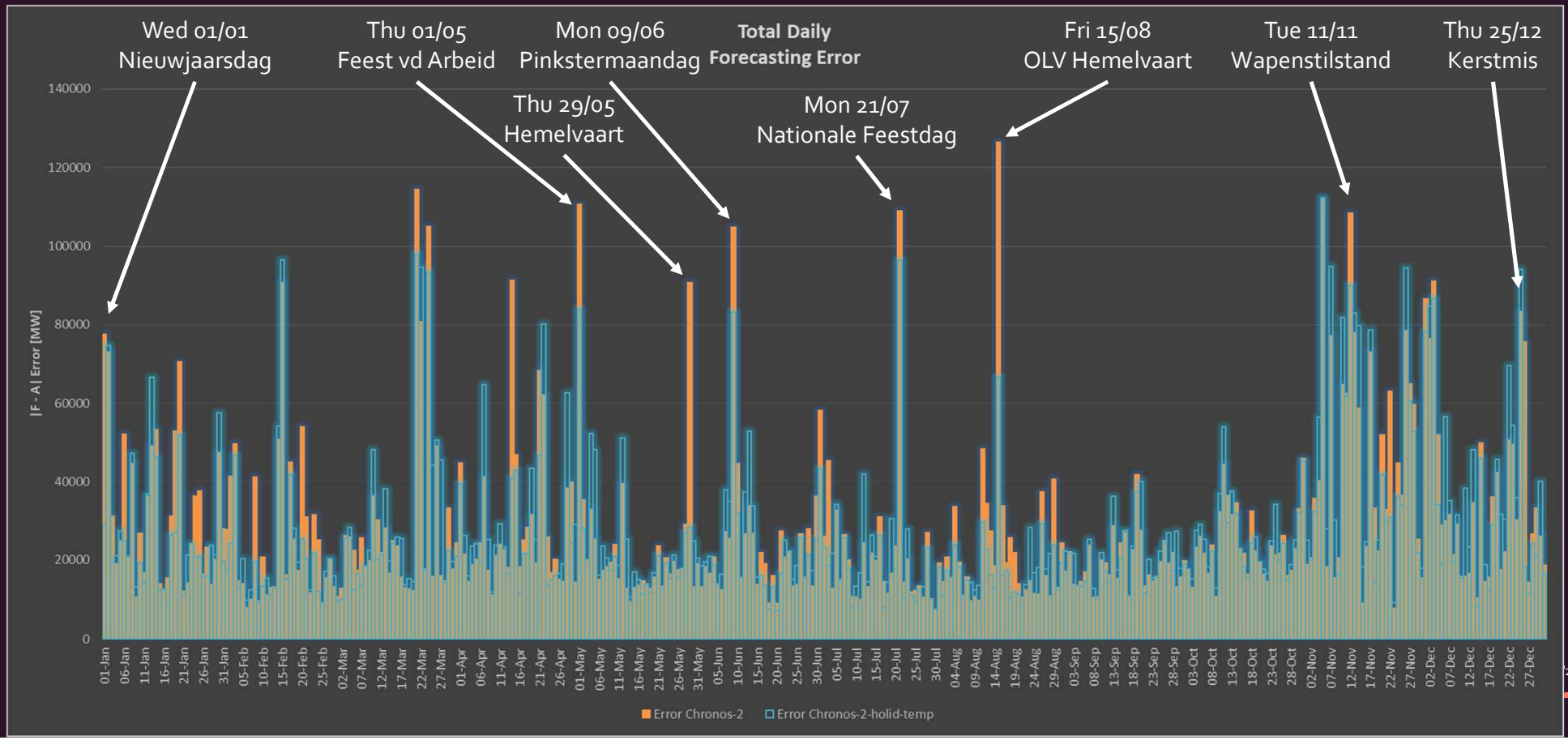


Adding holidays as covariates reduces those errors

BUT INCREASES ERRORS ON OTHER DAYS



Adding outdoor temperature reduces errors further

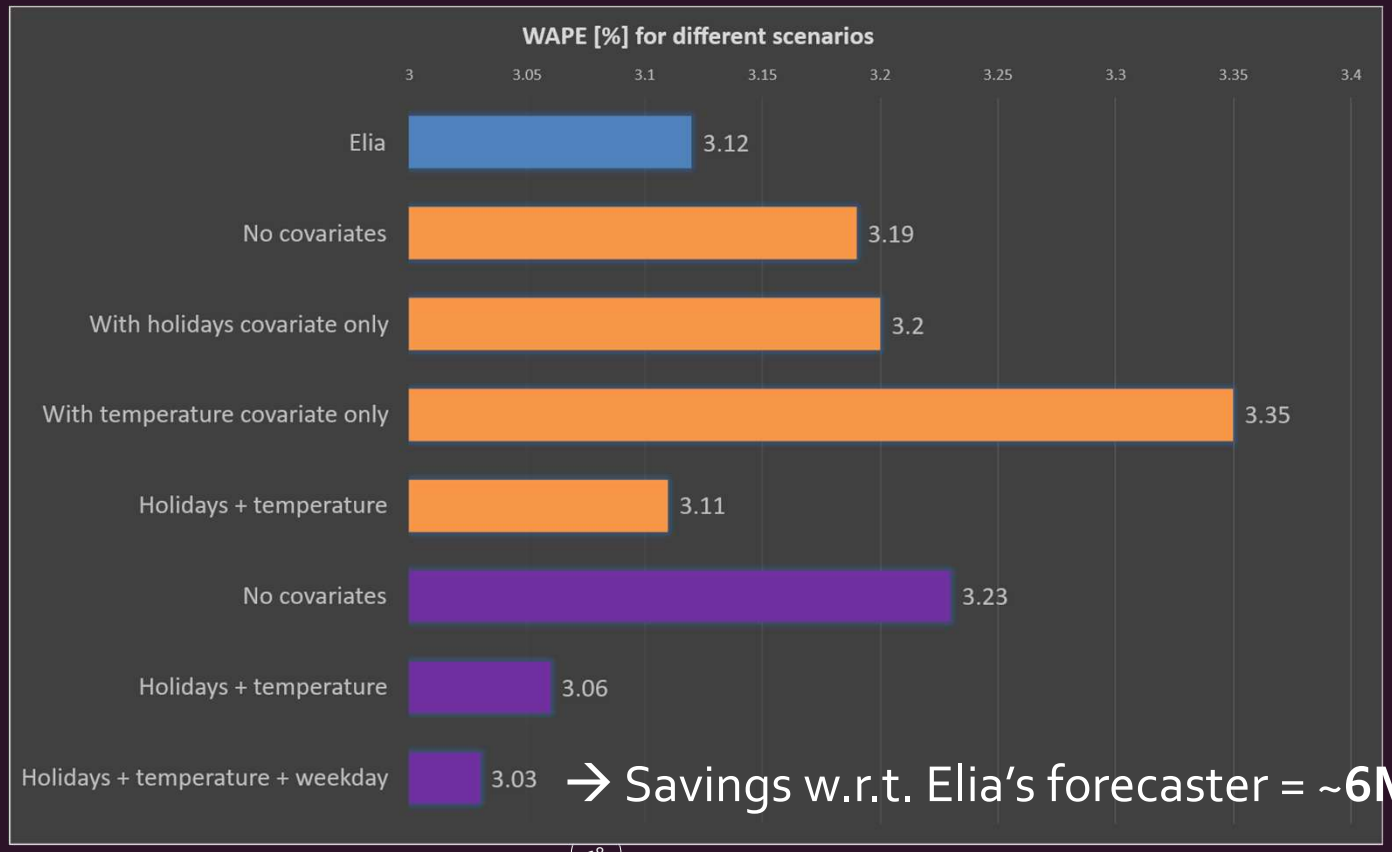


Chronos-2 with covariates achieves lowest error

LARGER CONTEXT WINDOW HELPS TOO

30 days context window

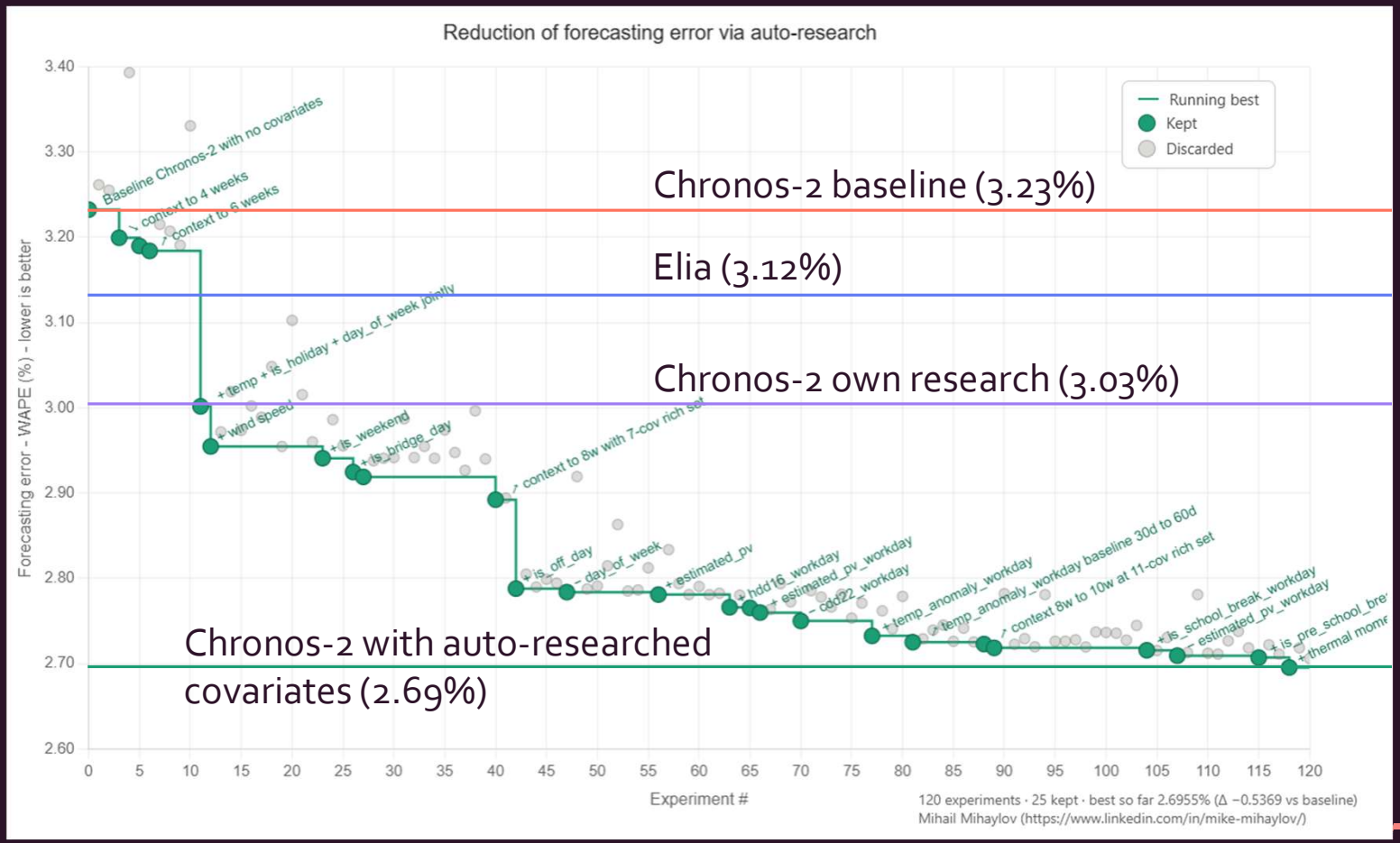
85 days context window[1]



→ Savings w.r.t. Elia's forecaster = ~**6M€/year**

[1] max context window for quarterly data: 85 days × 96 slots = 8160 time steps (max for Chronos-2 is 8192 time steps). Got same results with context window of 50 days.

“Auto-research” lowered error further



Use case 2: DAM price forecaster

DAY-AHEAD MARKET (DAM) PRICE OF ENERGY ON THE BELGIAN SPOT MARKET (BELPEX)

Company "M" offers 5 DAM price forecasting SaaS products

- Combine neural networks and evolutionary algorithms
- Use historical prices and 3 weather models
- Re-trained each day

I requested an offer as a not-for-profit org. → 6000€/month

Compared with Chronos-2 on 1y of hourly DAM prices. D+1 forecasts at 11am on day D.

Accuracy is within range. Used 80 days historical hourly DAM price. No training. Free.

	Company "M"s DAM price forecasting products					Chronos-2	
	"AI"	**"EQ"**	**"D"**	**"RAW"**	**"Finetuned"**	**Out-of-the-box**	**with Temp.**
WAPE	**15.2%**	17.4%	21.99%	22.41	25.03%	24.15%	23.31%

Chronos-2 surpasses latest literature benchmarks

A 2025 scientific paper[1] compares GenAI and classical ML models on diverse datasets.
I contacted authors and reproduced the results of the winning model across 4 datasets.
Compared Chronos-2 against benchmarks. **Achieved better accuracy** on all 4 of them.

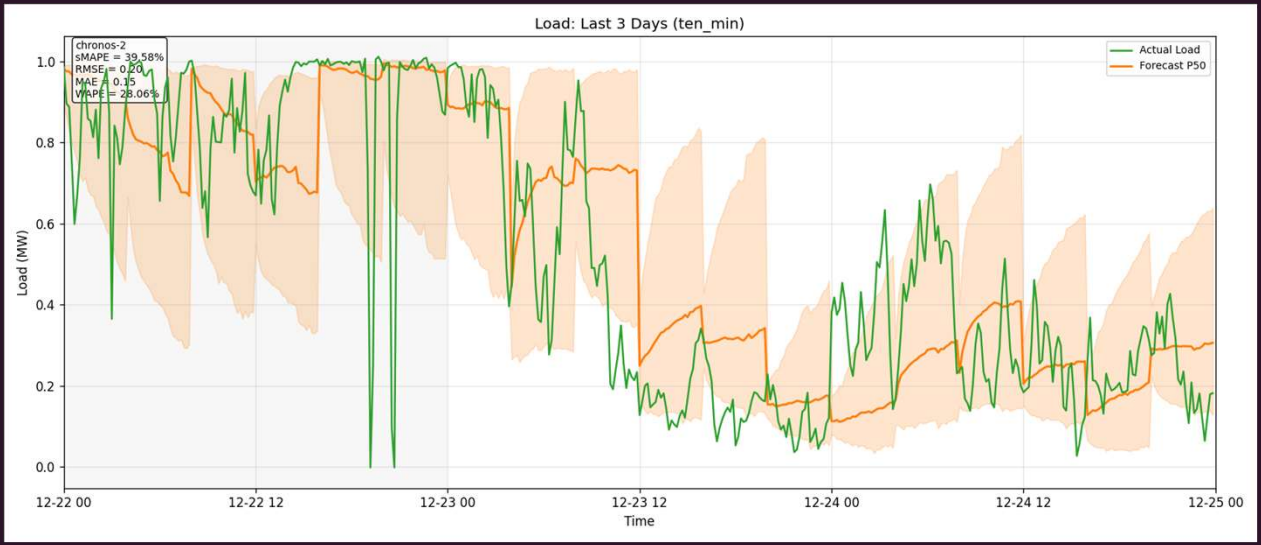
Dataset	Length	Granularity	WAPE Benchmark	Chronos-2
Oil Temperature v1 [°C]	2 years	Hour	20.57% (N-BEATS)	**11.52%**
Oil Temperature v2 [°C]	2 years	15 minutes	8.87% (PatchTST)	**8.78%**
Atmospheric pressure [mmHg]	137 days	10 minutes	0.33% (PatchTST)	**0.28%**
Seoul Bike Demand [no unit]	1 year	Hour	40.77% (N-BEATS)	**27.68%**

[1] Mojtaba A., et al. "Time-series forecasting in smart manufacturing systems: An experimental evaluation of the state-of-the-art algorithms", in Robotics and Computer-Integrated Manufacturing (95), 2025

Not a silver bullet for all use cases

Wind power generation of a single wind turbine – too much variability, lack of underlying pattern.

Forecasts made 4 hours ahead. Univariate. No covariates.



TFMs Are a Promising New Technology

Timeseries foundation models are disrupting the forecasting domain:

- As accurate as ML algorithms without the heavy engineering work
- General-purpose, cross-domain without re-training
- Free to use, even commercially

- Lower the cost needed for solving forecasting use cases
- Are here to stay
- Are the worst they will ever be

Thank you

Mihail Mihaylov
https://www.linkedin.com/in/mike-mihaylov/
mihail.mihaylov@sirris.be

innovation
forward