PROBABILISTIC FORECASTING VIA GENERATIVE AI

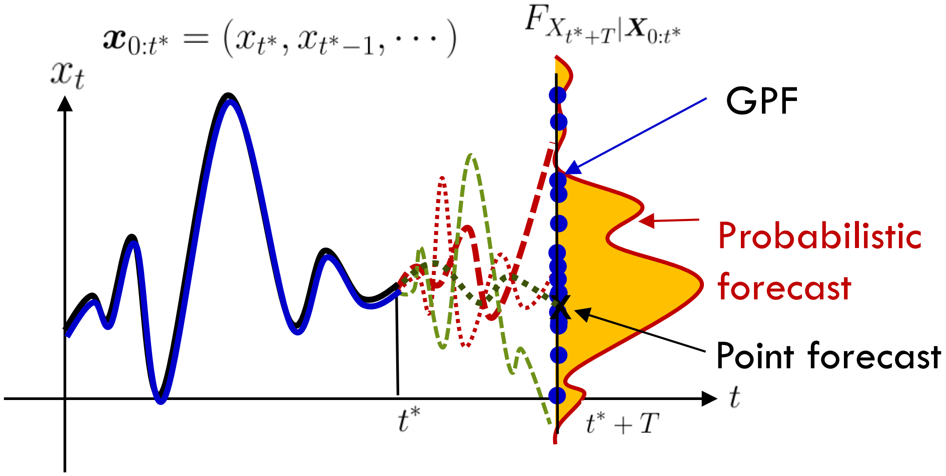
Lang Tong

Electrical and Computer Engineering
Cornell University, Ithaca, NY

Joint work with Qing Zhao and Kyla Wang

Presented at 7th AES Workshop, Golden, CO, September 5, 2024

Generative Probabilistic Forecasting (GPF)



State-of-art approaches

- **Parametric methods**

  Finite dimensional estimation and tracking

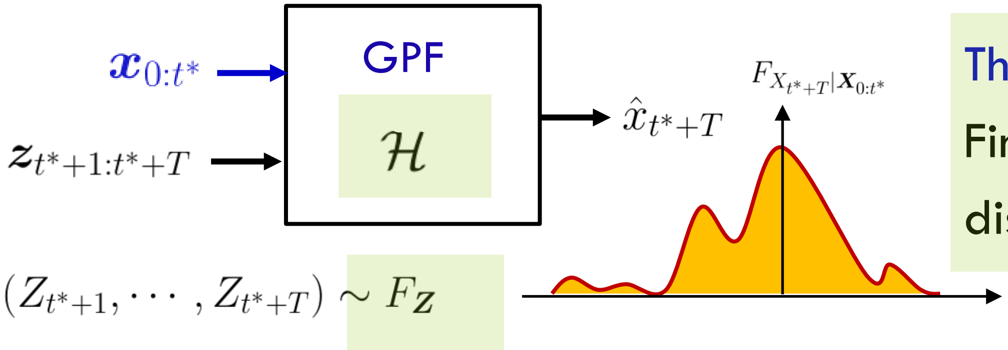
- **Non-parametric methods**

  Intractable. Infinite dimensional with few samples

- **Deep learning**

  A few, including using LMM and ChatGPT, none shown to solve the GPF problem.

A generic GPF structure



The GPF Problem

Find mapping $\mathcal{H}$ and sampling probability distribution $F_{\boldsymbol{Z}}$ such that $\hat{X}_{t^*+T} \sim F_{X_{t^*+T}|\boldsymbol{X}_{0:t^*}}$

Innovation Representation: Wiener-Kallianpur ('58) and Rosenblatt ('59)


Wiener


Kallianpur


Rosenblatt

- Wiener-Kallianpur's innovation hypothesis

$$(\cdots, x_{-2}, x_{-1}, x_0, x_1, \cdots, x_2, \cdots)$$

- Wiener-Kallianpur's strong innovation autoencoder

$$\nu_t = G(x_t, x_{t-1}, \cdots) \quad \hat{x}_t = H(\nu_t, \nu_{t-1}, \cdots) \overset{\text{a.s.}}{=} x_t$$

where $\nu_t \overset{i.i.d.}{\sim} U(0,1)$ and referred to as innovations.

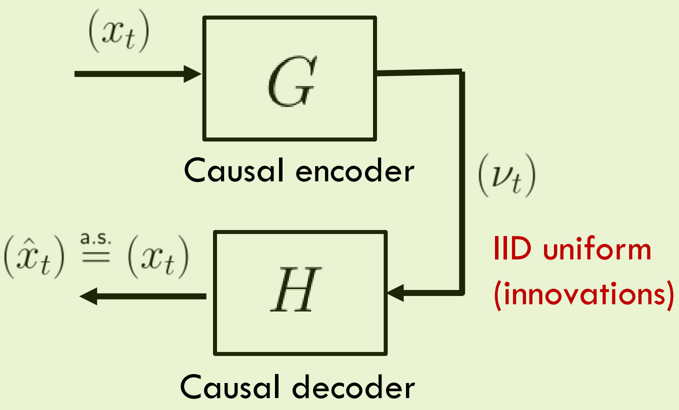
$$\nu_t = G(\overbrace{x_t, x_{t-1}, \cdots}^{\mathcal{X}_t}) = \tilde{G}(x_t, \overbrace{\nu_{t-1}, \nu_{t-2}, \cdots}^{\mathcal{X}_{t-1}})$$

$$x_{t-1}, \cdots = H(\nu_{t-1}, \nu_{t-1} \cdots) \Rightarrow \nu_t \perp\!\!\!\perp \mathcal{X}_{t-1}$$

contains only new information

Independent of the past!

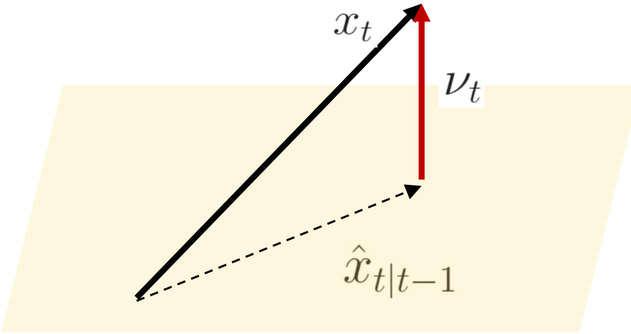
Innovation autoencoder (IAE)



- Rosenblatt's weak innovation autoencoder (WIAE): $(\hat{x}_t) \overset{\text{D}}{=} (x_t)$

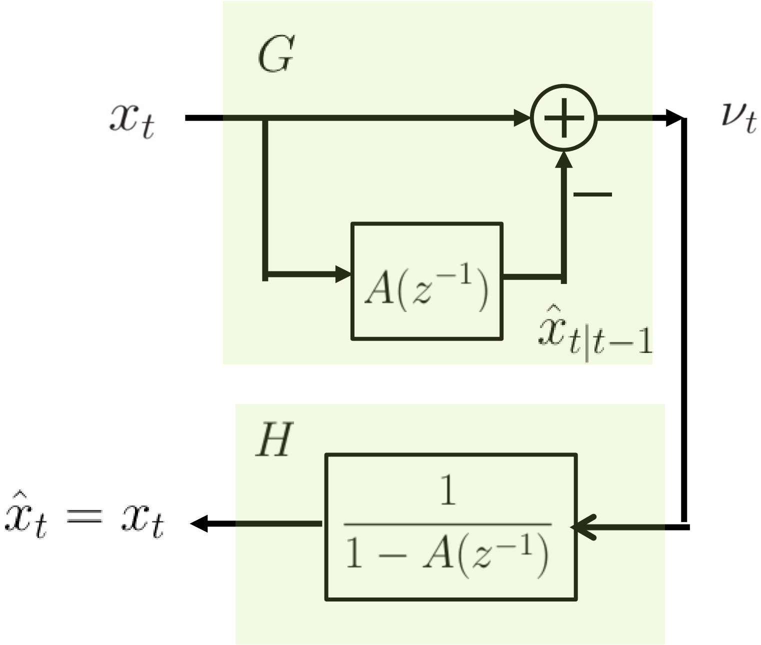
Computation of innovation autoencoder

- Gaussian model with linear minimum-mean-squared-error (MMSE) prediction (Kolmogorov, Wiener, Kalman)



$$\hat{x}_{t|t-1} := a_1 x_{t-1} + a_2 x_{t-2} + \cdots$$

$$\nu_t = x_t - \hat{x}_{t|t-1}.$$



- AWGN model in continuous-time:

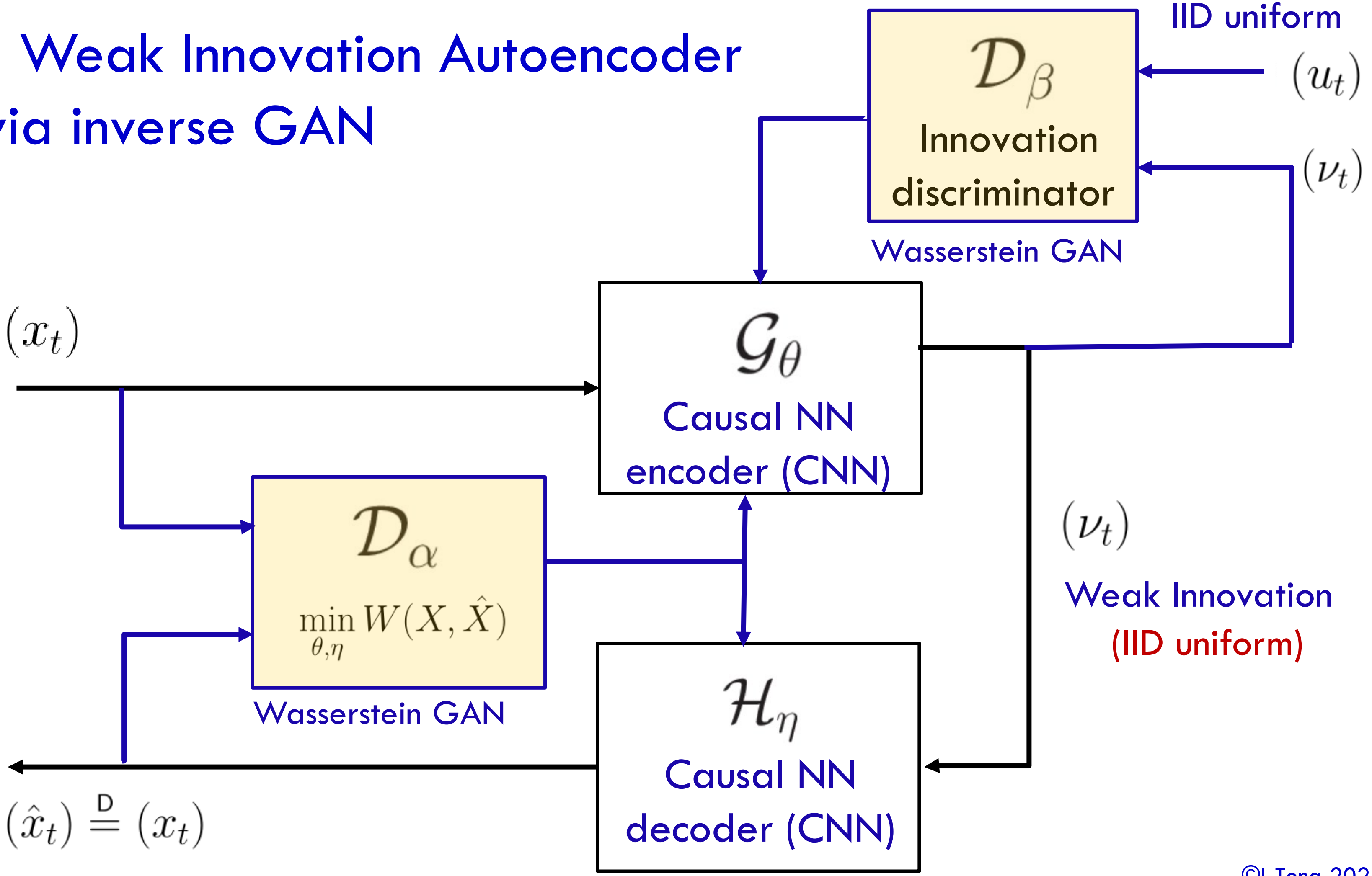
$$x(t) = s(t) + n(t)$$

Non-Gaussian Gaussian

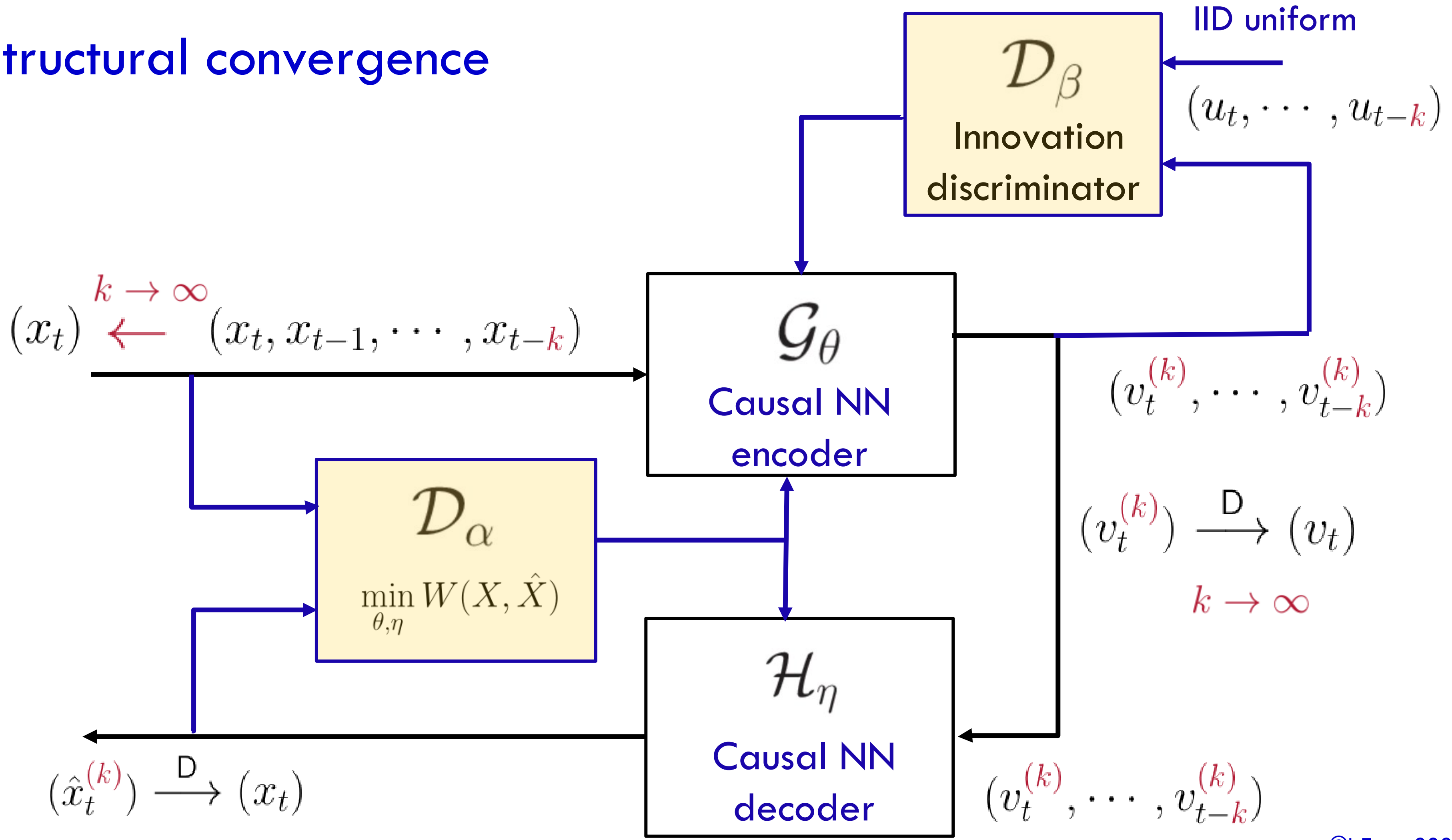
Nonlinear MSE prediction (Frost-Kailath)

$$\nu(t) = x(t) - \hat{x}(t)$$ Causal (nonlinear) minimum-mean-squared-error prediction

Learning Weak Innovation Autoencoder (WIAE) via inverse GAN

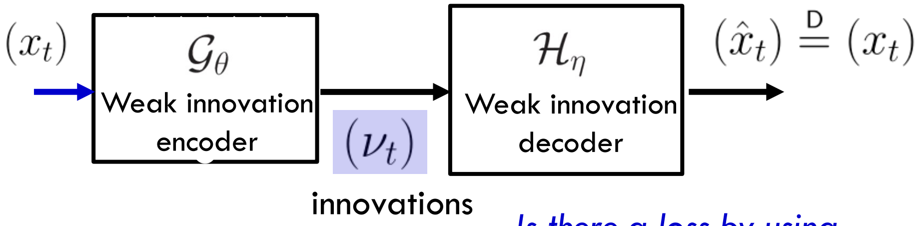


Structural convergence



GPF with Weak Innovation Auto-Encoder (WIAE)

WIAE

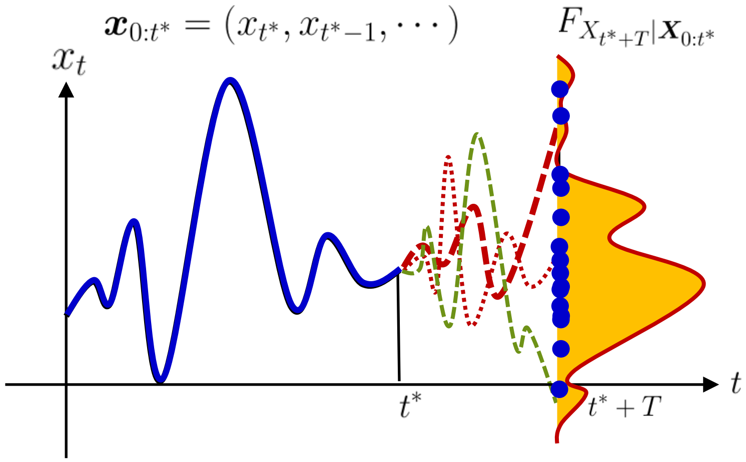


Is there a loss by using innovations for forecasting?

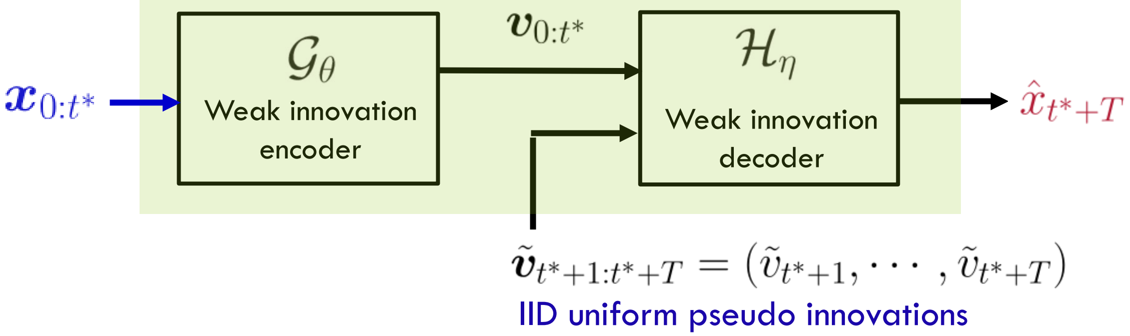
Lemma (Bayesian sufficiency)

$$F_{X_{t^*+T}|\boldsymbol{X}_{0:t^*}} = F_{X_{t^*+T}|\boldsymbol{V}_{0:t^*}}$$

"Optimal decisions based on weak innovations are lossless"



WIAE-GPF



Theorem (Validity) $F_{X_{t^*+T}|\boldsymbol{X}_{0:t^*}} = F_{\hat{X}_{t^*+T}|\boldsymbol{X}_{0:t^*}}$

Baselines and performance metrics

Algorithm	Forecasting Type	Time Series Model	Forecastor Output	ML Models
SNARX [4]	Probabilistic	Semiparametric AR	AR Model Parameters	Kernel Estimation
WIAE-GPF	Probabilistic	Nonparametric	Generative	CNN + WIAE
TLAE [17]	Probabilistic	Parametric	Generative	RNN + VAE
DeepVAR [36]	Probabilistic	Parametric (AR Model)	Model Parameters	LSTM
BWGVT [16]	Probabilistic	Nonparametric	Forecasted Quantiles	LLM + Quantile Regression
Pyraformer [28]	Point	Nonparametric	Point Estimate	LLM
Informer [27]	Point	Nonparametric	Point Estimate	LLM

$$\text{NMSE} = \frac{\frac{1}{N}\sum_{t=1}^{N}(\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t)^2}{\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{X}_t^2},$$

$$\text{NMAE} = \frac{\frac{1}{N}\sum_{t=1}^{N}|\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t|}{\frac{1}{N}\sum_{t=1}^{N}|\boldsymbol{X}_t|},$$

$$\text{MASE} = \frac{\frac{1}{N}\sum_{t=1}^{N}|\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t|}{\frac{1}{N-T}\sum_{t=T+1}^{N}|\boldsymbol{X}_t - \boldsymbol{X}_{t-T}|},$$

Mean Absolute Scaled Error

$$\text{sMAPE} = \frac{1}{N}\sum_{t=1}^{N}\frac{|\boldsymbol{X}_t - \tilde{\boldsymbol{X}}_t|}{(|\boldsymbol{X}_t| + |\tilde{\boldsymbol{X}}_t|)/2}.$$

Absolute % error

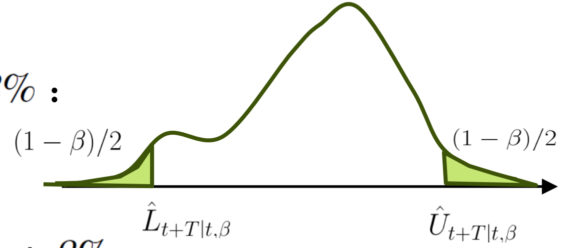
Continuously ranked probability score:

Prob Forecast Realized

$$\text{CRPS} = \int \tilde{\mathbb{E}}\left[\tilde{F}_{t+T|t}(\boldsymbol{x}|\boldsymbol{x}_{0:t}) - \mathbb{1}_{\{\boldsymbol{x}_{t+T}\le \boldsymbol{x}\}}\right]^2 d\boldsymbol{x}$$

Coverage Probability Error (CPE) at $\beta\%$:

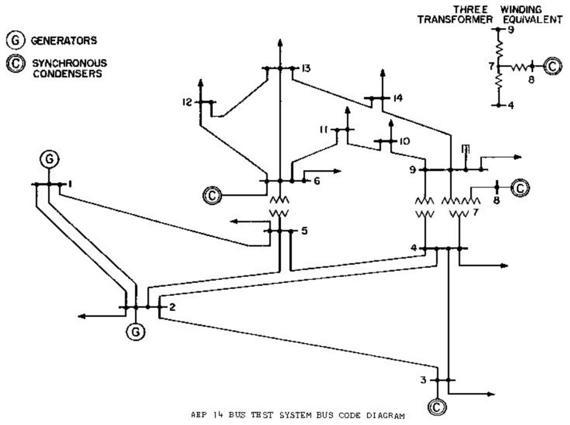
$$\text{CPE}(\beta\%) = \text{CP}(\beta\%) - \beta\%.$$



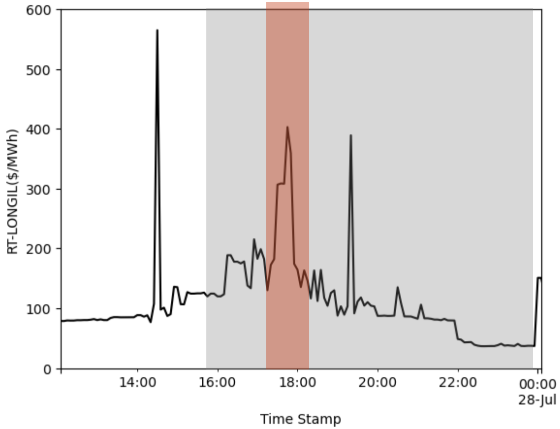
Normalized Coverage Width (NCW) at $\beta\%$:

$$\text{NCW}(\beta\%) = \frac{1}{N}\sum_{t=1}^{N-T}\frac{\hat{U}_{t+T|t,\beta} - \hat{L}_{t+T|t,\beta}}{\hat{U}_\beta - \hat{U}_\beta}$$

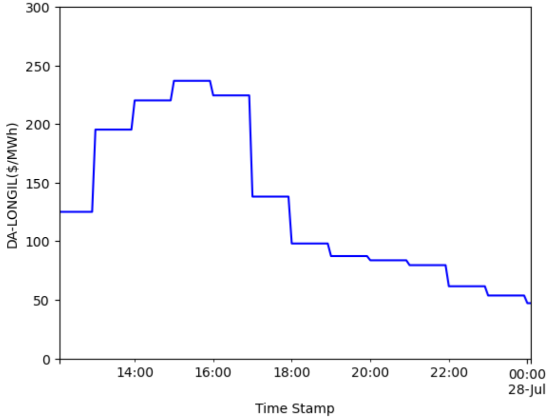
Real-time LMP forecasting



RT LMP @ LONGIL



DA LMP @ LONGIL



DC-OPF (shift-factor form)

minimize $C(\boldsymbol{p}) = \sum_{i=1}^{n} C_i(p_i)$

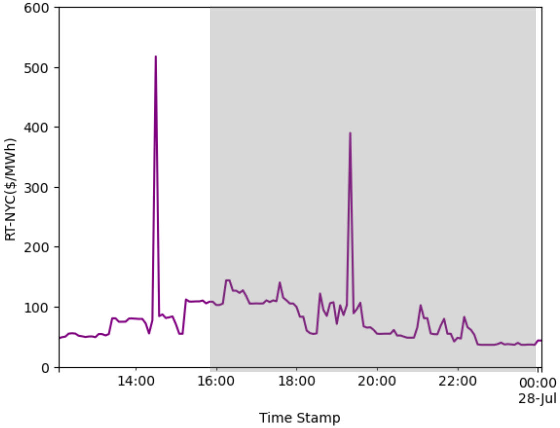
subject to $\underline{\boldsymbol{p}} \leq \boldsymbol{p} \leq \overline{\boldsymbol{p}}$

(Power balance) $\mathbb{1}^\top \boldsymbol{p} = \mathbb{1}^\top \boldsymbol{d}$ (λ)

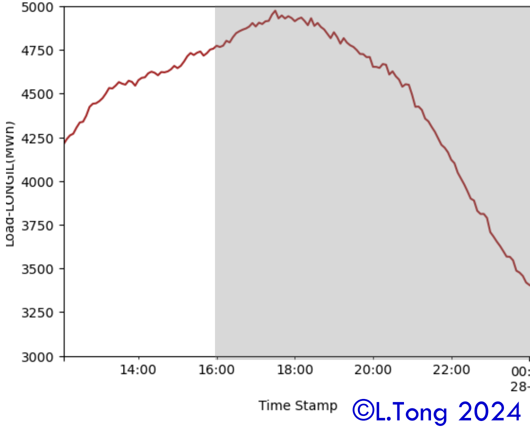
(Congestion) $\boldsymbol{S}(\boldsymbol{p} - \boldsymbol{d}) \leq \boldsymbol{f}$ $(\boldsymbol{\mu})$

LMP: $\boldsymbol{\pi}^{\mathrm{LMP}} = \lambda^* \mathbb{1} + \tilde{\boldsymbol{S}}^\top \boldsymbol{\mu}^*$

RT LMP @ NYC



RT load @ LONGIL

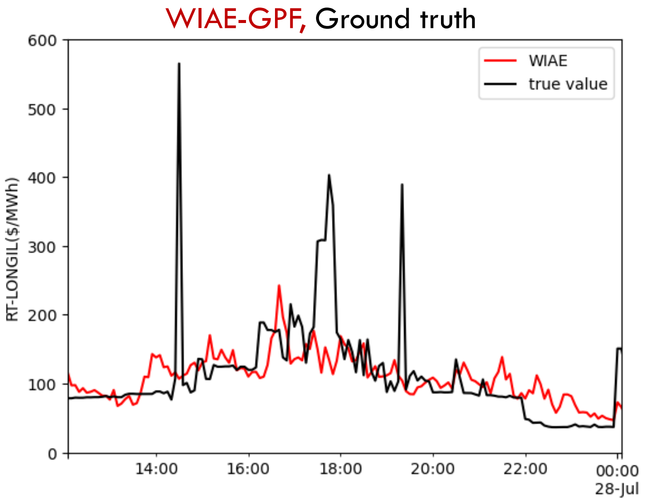


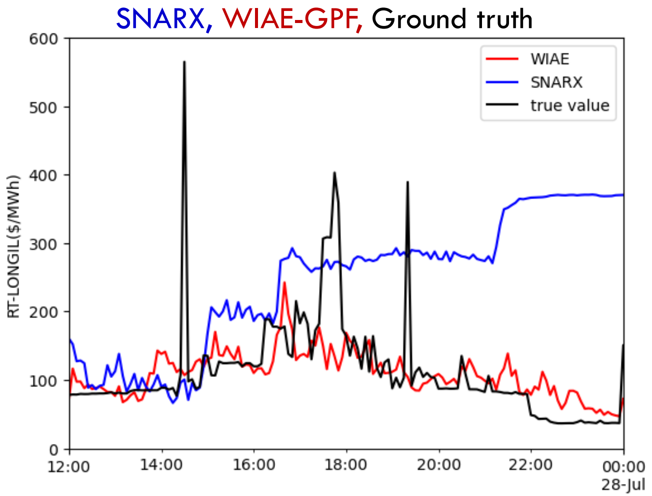
LMP forecasting at LONGIL of NYISO

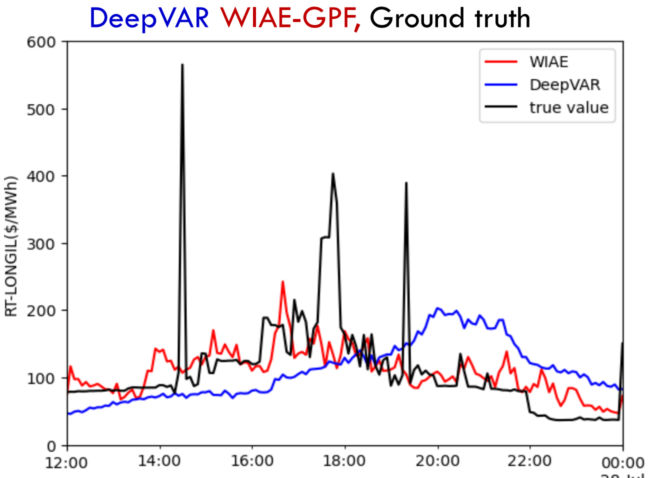
	Methods	NMSE		NMAE	
		LONGIL	LONGIL & NYC	LONGIL	LONGIL & NYC
Kernel Estimation	SNARX	0.9852	0.5029	0.9733	0.7318
CNN + WIAE	WIAE-GPF	**0.0585**	0.0487	**0.2074**	**0.1186**
RNN + VAE	TLAE	0.2956	**0.0232**	0.4186	0.1366
LSTM	DeepVAR	0.3919	0.4060	0.4088	0.4097
LLM	BWGVT	0.2670	0.2528	0.3158	0.3280
LLM	Pyraformer	0.3128	0.1382	0.3074	0.2556
LLM	Informer	0.1912	0.1829	0.3147	0.2830

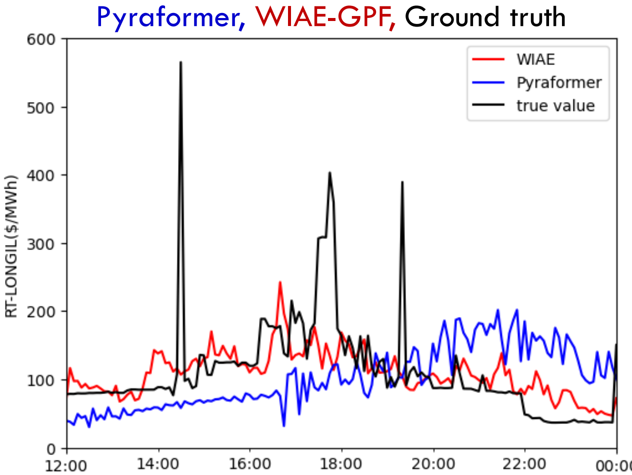
Methods	MASE		sMAPE		CRPS	
	LONGIL	LONGIL & NYC	LONGIL	LONGIL & NYC	LONGIL	LONGIL & NYC
SNARX	1.0666	0.6294	1.4319	0.4162	18.2864	10.5286
WIAE-GPF	**0.1503**	**0.0737**	**0.0839**	**0.0316**	**4.0029**	**1.1519**
TLAE	0.2968	0.0917	0.2720	0.3190	6.1875	2.8449
DeepVAR	0.2832	0.2841	1.3629	0.3703	23.1460	22.9434
BWGVT	0.8890	0.2979	0.2817	0.3900	24.2595	24.3065
Pyraformer	0.2334	0.1538	0.1059	0.3765	N/A	N/A
Informer	0.4786	0.1802	0.3761	0.3827	N/A	N/A

One-hour-ahead LMP forecasting with 5-hour history



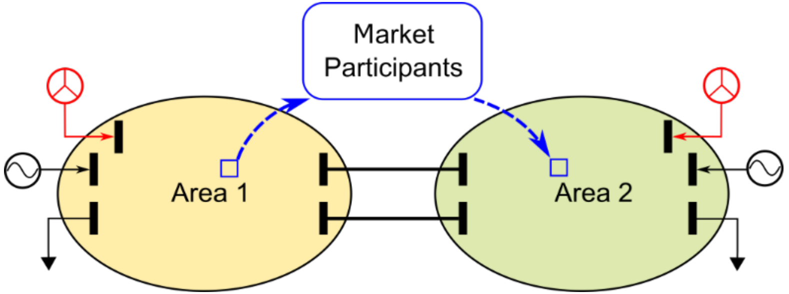






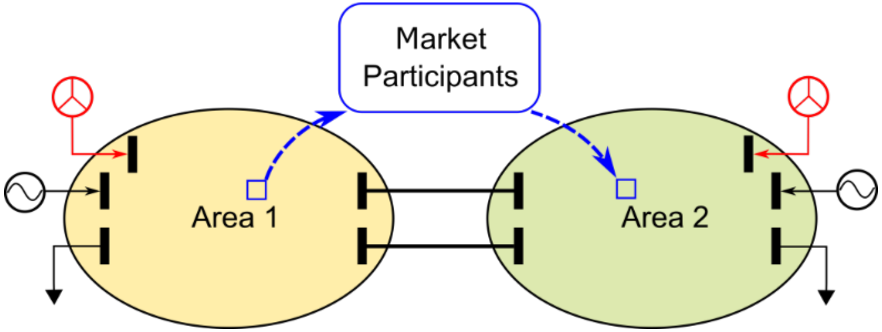
Interchange market with CTS

- Independent System Operators (ISOs) do not trade power directly with each other; market participants facilitate trades.
- Market participants submit (virtual) bids/offers at specific "proxy buses," buying Q at one proxy and selling Q at the other.



- ISOs clear these bids/offers (separately or jointly) and set the interchange quantity by coordinating cleared bids ahead of time.
- Market participants do not incur physical obligations to generate/consume.
- Cleared trades are financially binding; they are settled based on the real-time locational marginal price (LMP).
- A market participant is paid if it bets the price-spread direction correctly. Otherwise, it will have to pay to the operator.

NYISO-PJM Interchange



NYISO: Sandy Point   PJM: Neptune

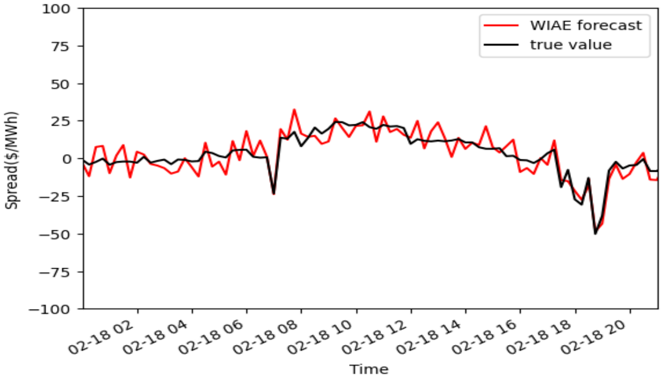
- Timing: Market closing 75 minutes before delivery
- Data: 15-minute LMP
- 75 min ahead forecasting with 5 hour past LMP and load

Probability of Correct Price Direction

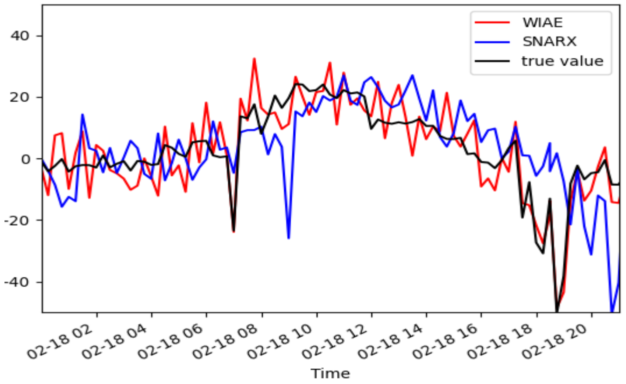
	Methods	NMSE	NMAE	MASE	sMAPE	PCPD	CRPS
Kernel Estimation	SNARX	2.4531	1.3415	1.5762	0.4958	0.3282	120.0403
CNN + WIAE	WIAE-GPF	**0.0098**	**0.2738**	**0.2418**	**0.4493**	**0.9394**	**4.0329**
RNN + VAE	TLAE	0.9592	0.9785	0.9516	0.4785	0.6308	15.5195
LSTM	DeepVAR	8.9864	0.7224	0.8253	0.4806	0.6495	32.8296
LLM	BWGVT	0.9053	0.8525	0.6513	0.4674	0.7687	31.5660
LLM	Pyraformer	0.9478	1.2674	0.9796	0.4909	0.3262	N/A
LLM	Informer	0.8045	0.4185	0.4836	0.4580	0.4513	N/A

75-min-ahead price spread forecasting @ NYISO-ISONE Interchange

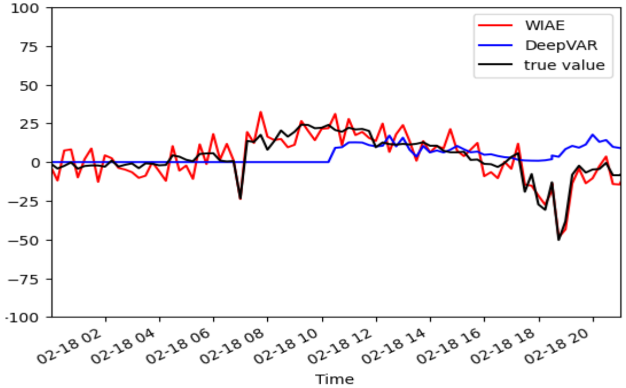
WIAE-GPF, Ground truth



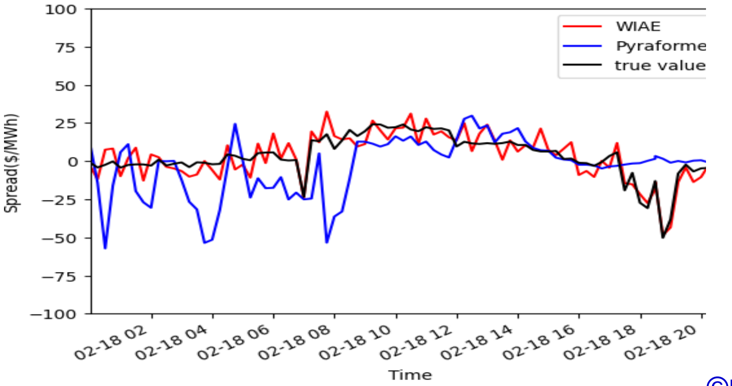
SNARX, WIAE-GPF, Ground truth



DeepVAR, WIAE-GPF, Ground truth

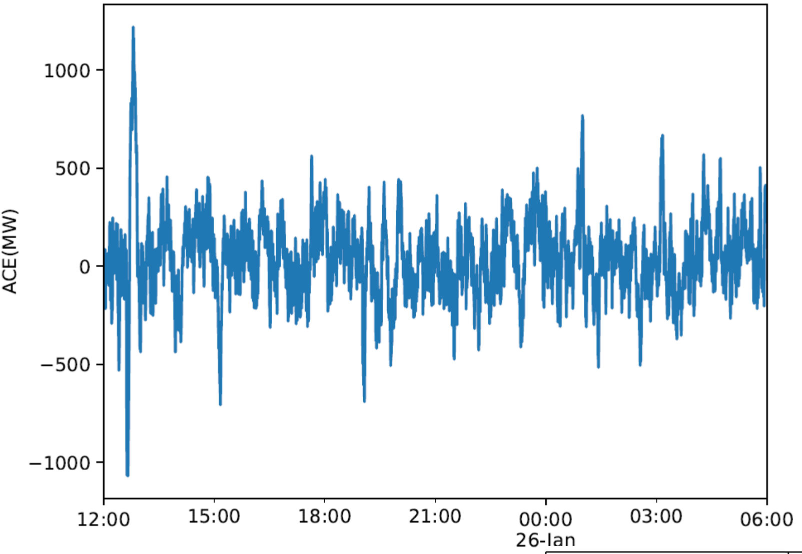


Pyraformer WIAE-GPF, Ground truth



Generative Probabilistic Forecasting of Area Control Error (ACE)

PJM ACE during Jan 25-26, 2023

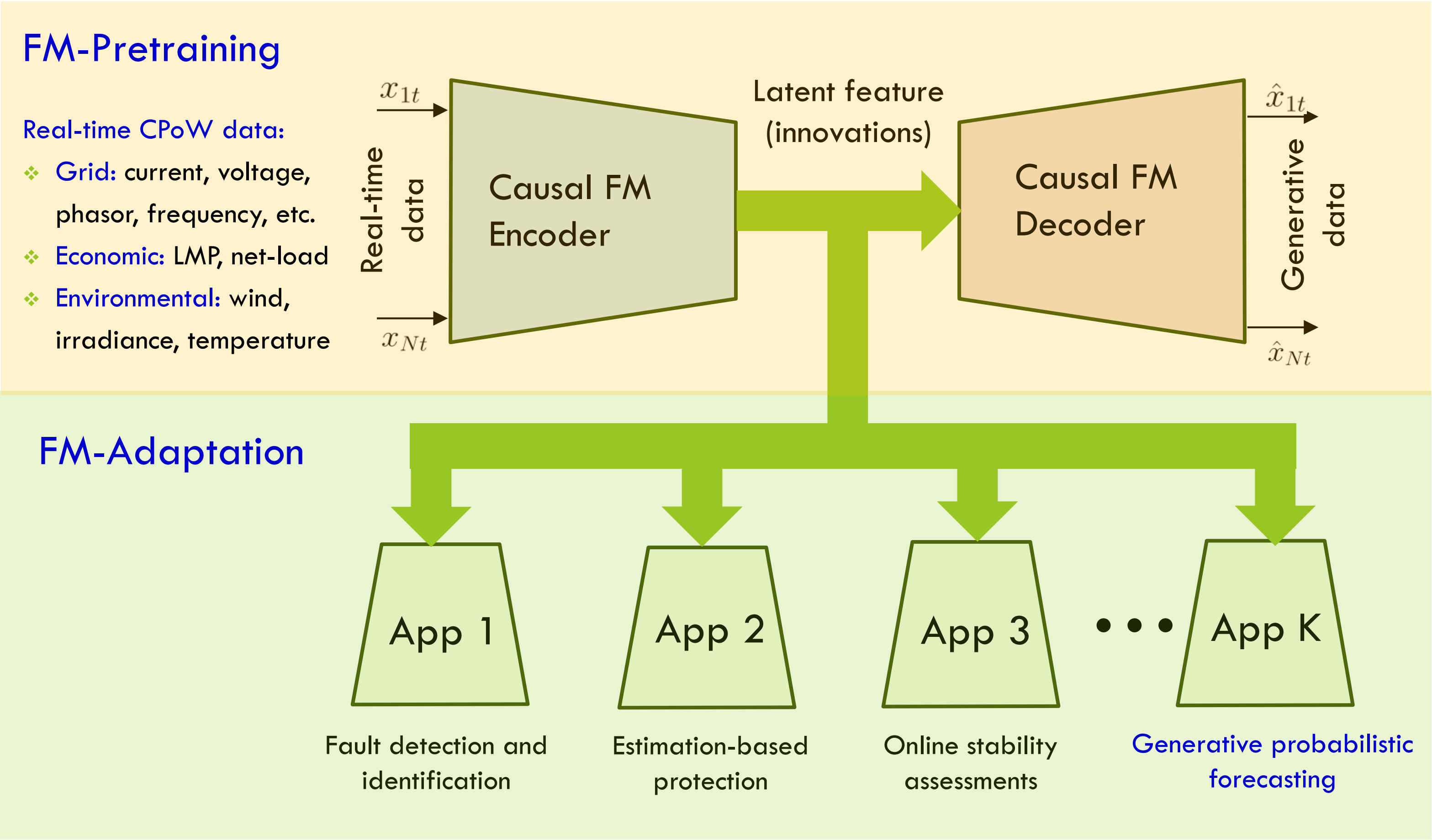


Methods	NMSE	NMAE	MASE	sMAPE
WIAE-GPF	(1) **0.5957**	(1) **0.7555**	(1) **0.4698**	(1) **0.1059**
TLAE [21]	(5) 1.1727	(5) 1.0605	(5) 0.6595	(3) 0.2782
DeepVAR [11]	(7) 1.4431	(7) 1.1750	(7) 0.7307	(5) 0.3952
BWGVT [19]	(3) 0.9562	(2) 0.9793	(2) 0.6090	(4) 0.3168
Pyraformer [34]	(4) 0.9783	(4) 0.9948	(4) 0.6186	(7) 0.4986
Informer [33]	(2) 0.6006	(3) 0.9819	(3) 0.6106	(2) 0.2247

- 4-sec timescale ACE.
- 5-min (200-step) ahead forecasting.

Methods	CRPS	CPE (90%) [NCW]	CPE (50%) [NCW]	CPE (10%) [NCW]
WIAE-GPF	(1) **0.0081**	(1) **−0.0016** [0.9199]	(1) **0.0321** [0.9336]	(1) **−0.0132** [0.8885]
TLAE [21]	(4) 1.5541	(4) −0.7857 [0.0004]	(4) −0.4489 [0.0005]	(4) −0.0957 [0.0027]
DeepVAR [11]	(3) 1.2947	(3) −0.3526 [0.5665]	(3) −0.2560 [0.5296]	(3) −0.0521 [0.5434]
BWGVT [19]	(2) 1.2488	(2) 0.0065 [1.8309]	(2) 0.0754 [2.0385]	(5) 0.0996 [2.4261]

A Foundation Model approach to system and market operations



Conclusion

- Learning probability distribution is very difficult, especially if it is learning a future (conditional) distribution. Yet, GPF is possible.

- The Wiener-Kallianpur-Rosenblatt *innovation autoencoder* is a canonical architecture for GPF and a Foundation Model for real-time non-parametric decision-making and control.

References

[1] Xinyi Wang and Lang Tong, Innovations Autoencoder and its Application in One-class Anomalous Sequence Detection, *Journal of Machine Learning Research*, vol. 23, no. 49., pp. 1-27, 2022.

[2] X. Wang, L. Tong and Q. Zhao, "Generative Probabilistic Time Series Forecasting and Applications in Grid Operations *2024 58th Annual Conference on Information Sciences and Systems (CISS)* Princeton, NJ, March, 2024.

[3] X. Wang, Q. Zhao, and L. Tong, "Forecasting Electricity Market Signals with Generative AI," arXiv:2403.05743.