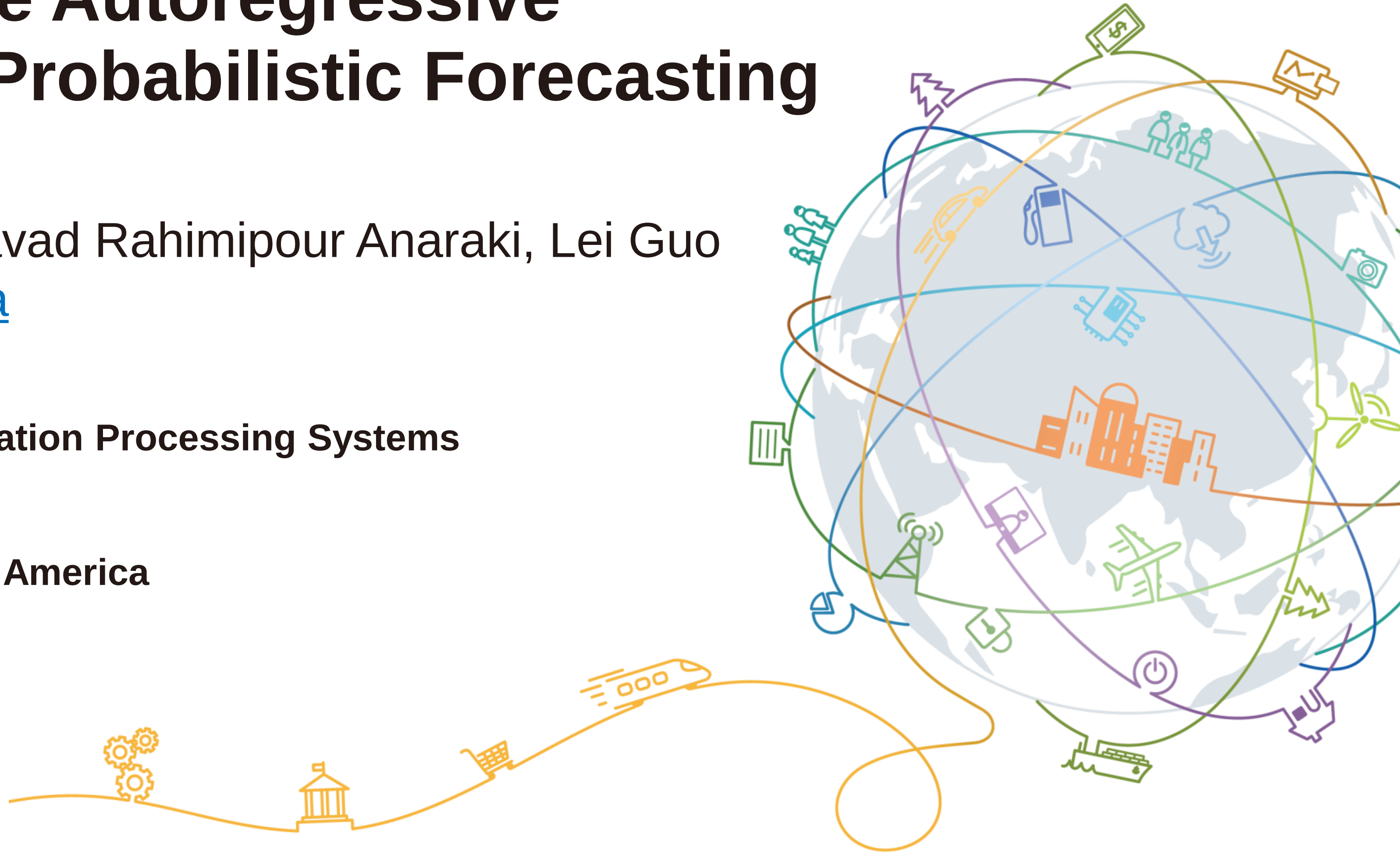
HUAWEI

C2FAR: Coarse-to-Fine Autoregressive Networks for Precise Probabilistic Forecasting

Shane Bergsma, Timothy Zeyl, Javad Rahimipour Anaraki, Lei Guo
Huawei Cloud, Alkaid Lab Canada

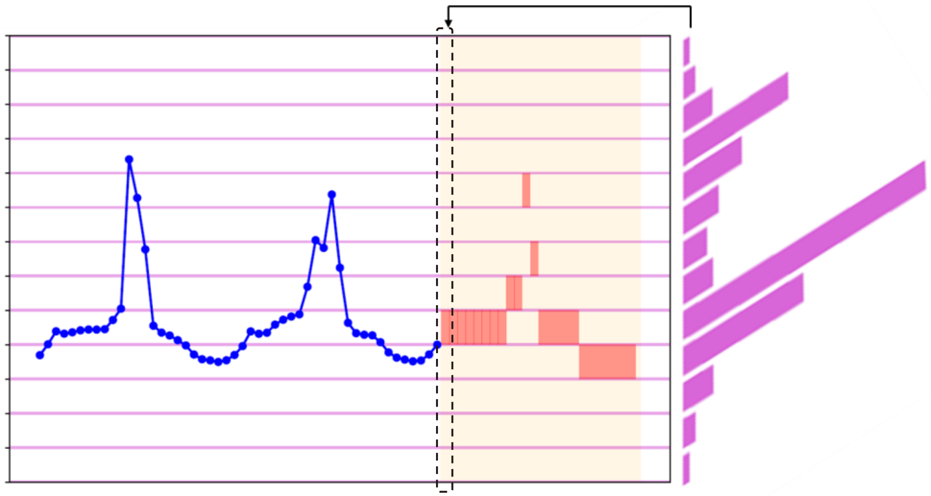
Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS 2022)
Nov. 28 – Dec. 9, 2022
New Orleans, Louisiana, United States of America



C2FAR: Overview

Prior generative models for numeric data:

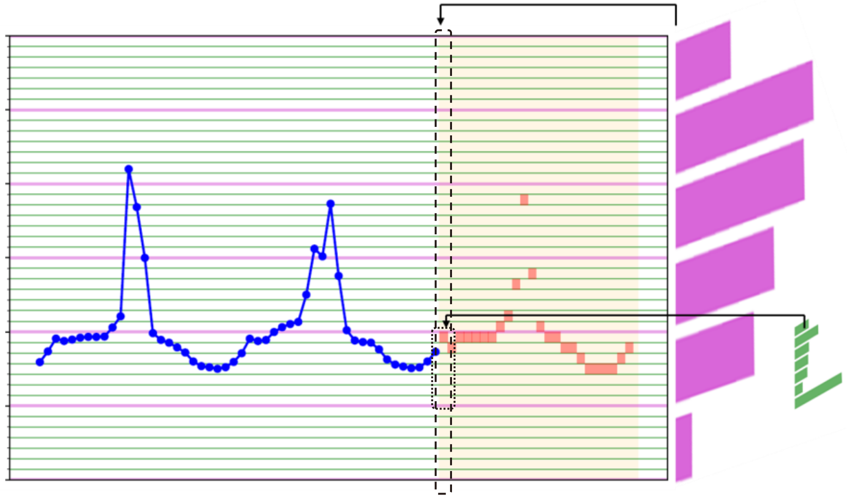
- use parametric density, e.g. negative binomial (for discrete data), Gaussian (for continuous), or
- bin (discretize) data, use a softmax distribution:



Problem: more precision ➔ more bins ➔ more parameters

C2FAR (Coarse-to-Fine Auto-Regressive) networks:

- bin (discretize) data *hierarchically*, first generating a coarse bin (from coarse softmax), then a finer bin conditional on the coarse bin, etc., autoregressively



C2FAR: Exponentially more precision ➔ linearly more parameters

- C2FAR networks implicitly define a piecewise-uniform continuous density; we use special Pareto densities in (unbounded) extreme high/low bins, enabling handling of *unbounded* (infinite-scale) data
- We use C2FAR within a RNN for **probabilistic forecasting**, achieving state-of-the-art accuracy

Related Work

Continuous data:

- Binned forecasting models (Rabanser et al., 2020) with “spliced” Pareto tails (Ehrlich et al., 2021)

Discrete data:

- Multi-stage likelihoods with zero-inflation (Seeger et al., 2016)
- WaveRNN’s dual softmax (Kalchbrenner et al., 2018)

Non-numeric data:

- Hierarchical softmaxes in language modeling (Morin & Bengio, 2005)

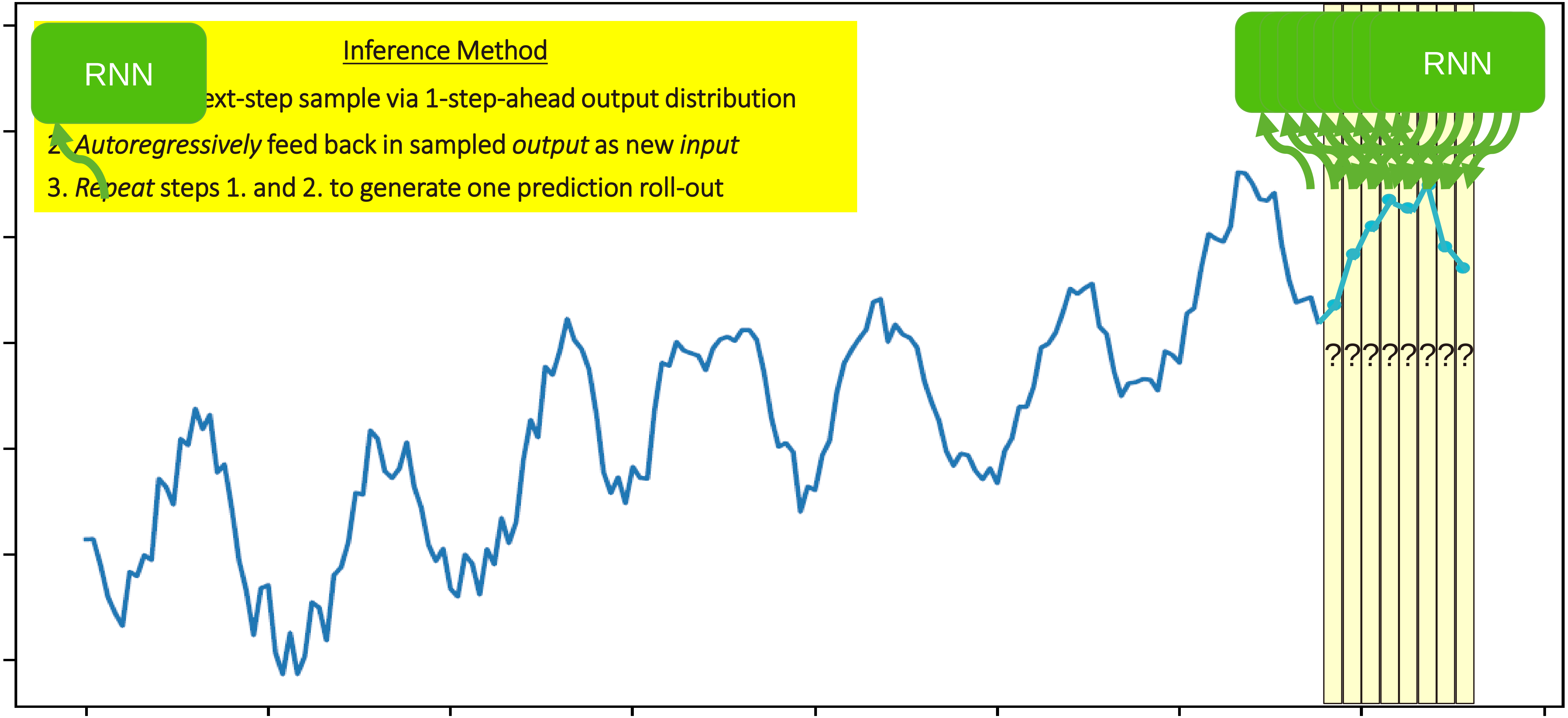
C2FAR generalizes and unifies prior work in modeling numeric data.

Autoregressive Probabilistic Forecasting (DeepAR)

Salinas et al., 2020

- Train a global sequence model on a dataset of related time series.

Autoregressive Probabilistic Forecasting (DeepAR)

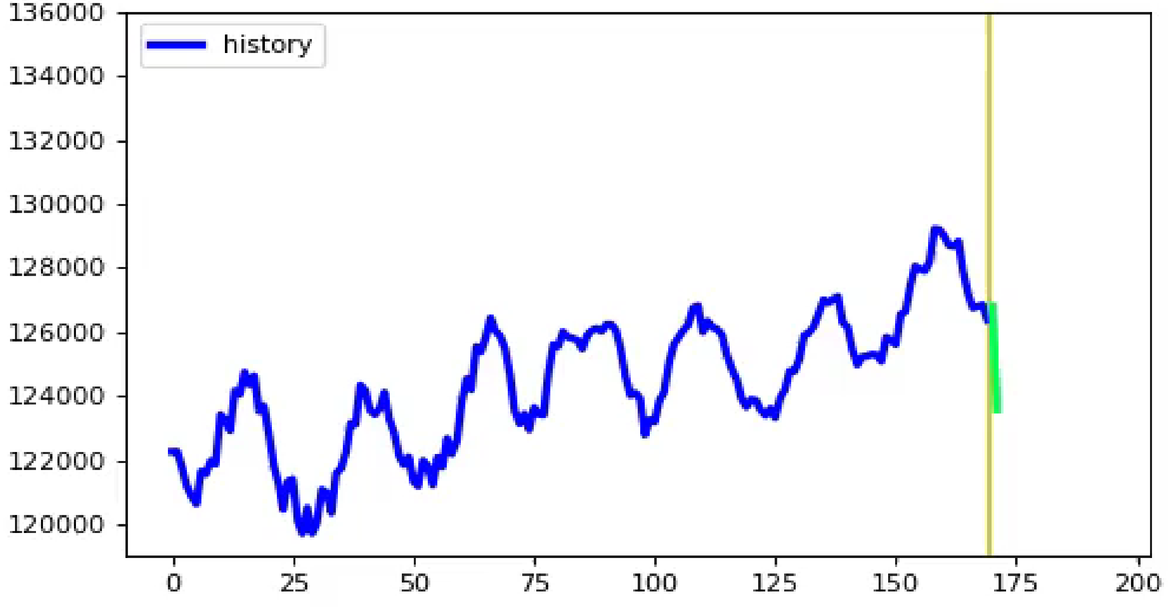


HUAWEI

Autoregressive Probabilistic Forecasting (DeepAR)

Generate many rollouts to obtain Monte Carlo estimate of full joint probability of future:

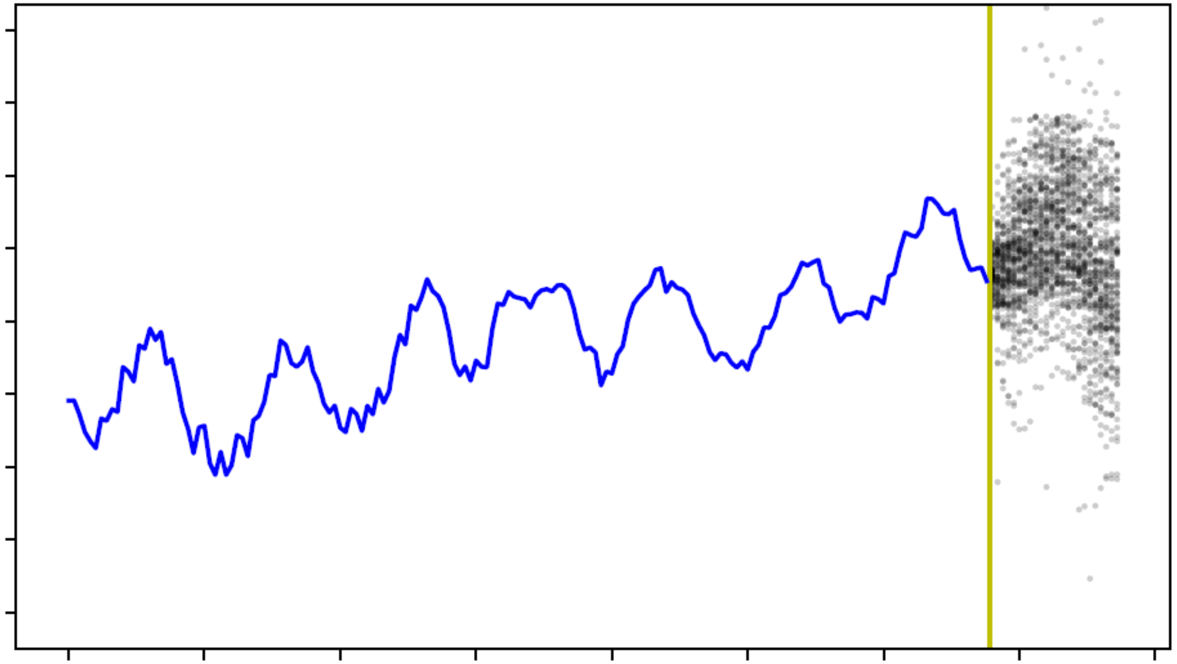
$$p(z_{T+1} \ldots z_{T+N} | z_1 \ldots z_T, \mathbf{x}_1 \ldots \mathbf{x}_{T+N})$$

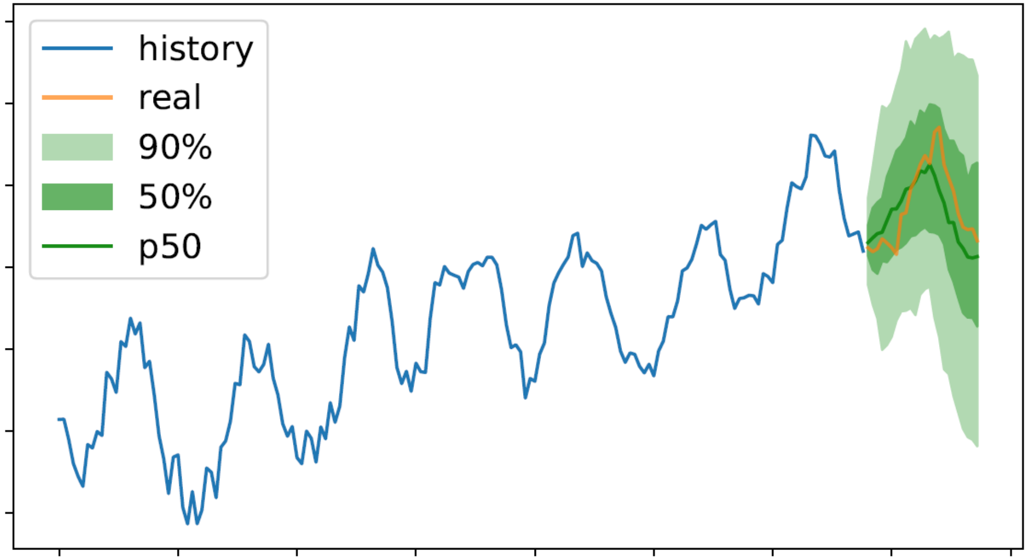


Autoregressive Probabilistic Forecasting (DeepAR)

Generate many rollouts to obtain Monte Carlo estimate of full joint probability of future:

$$p(z_{T+1} \ldots z_{T+N} | z_1 \ldots z_T, \mathbf{x}_1 \ldots \mathbf{x}_{T+N})$$

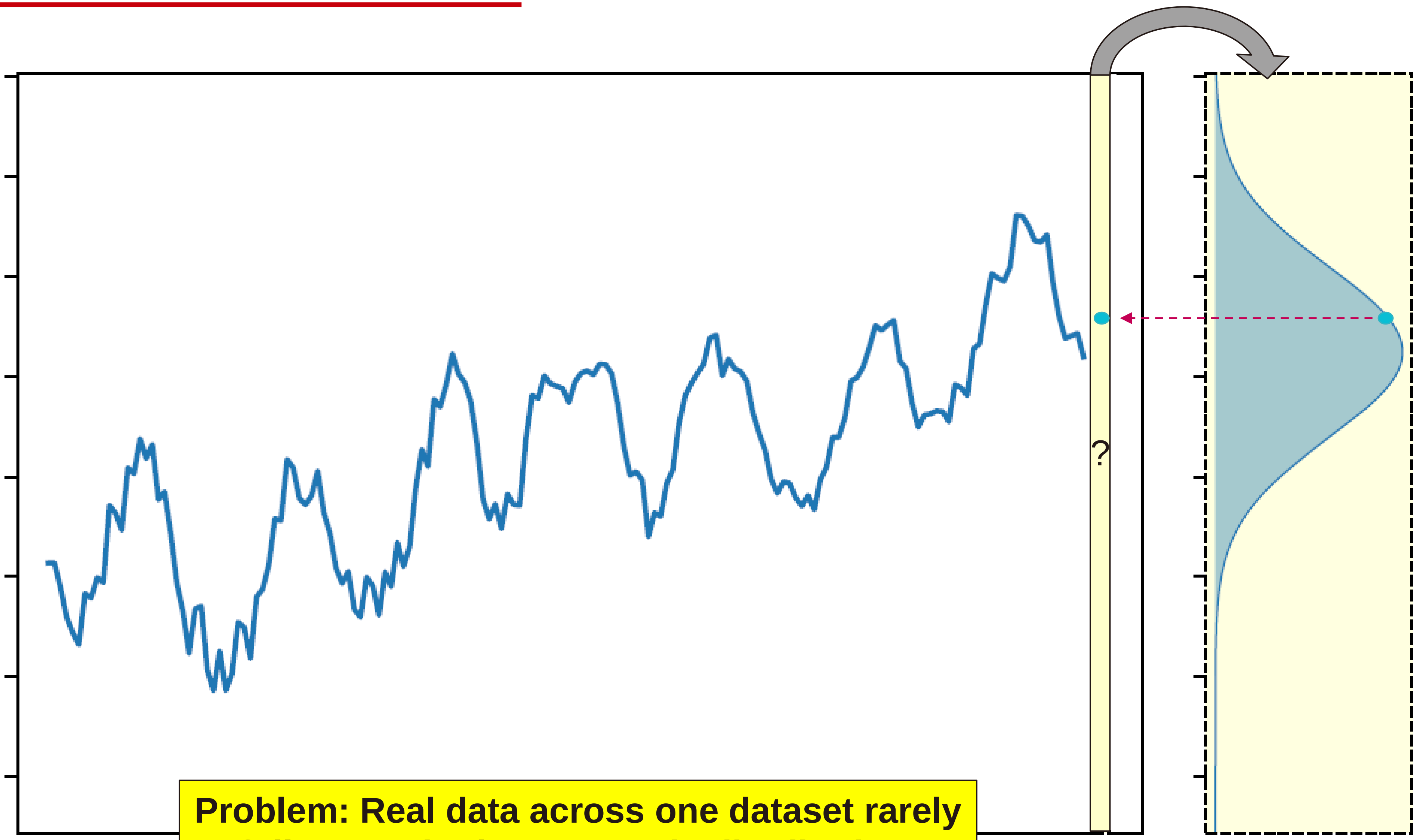




From which we can get desired forecast quantiles for downstream decision making.

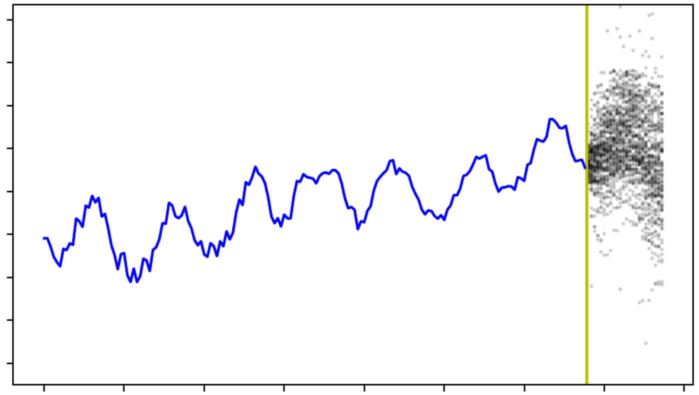
HUAWEI

Parametric Densities

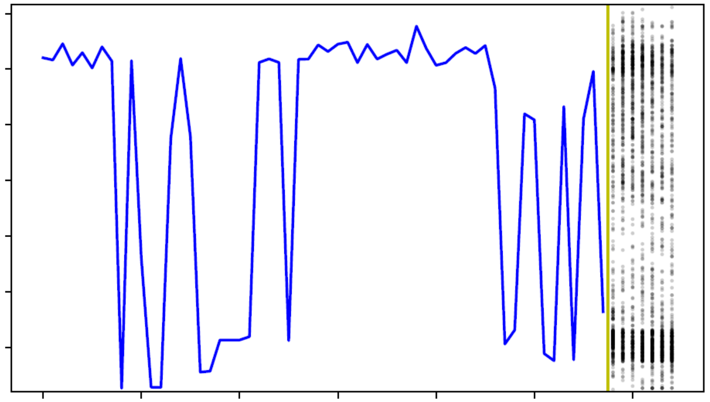


Problem: Real data across one dataset rarely follows a single parametric distribution.

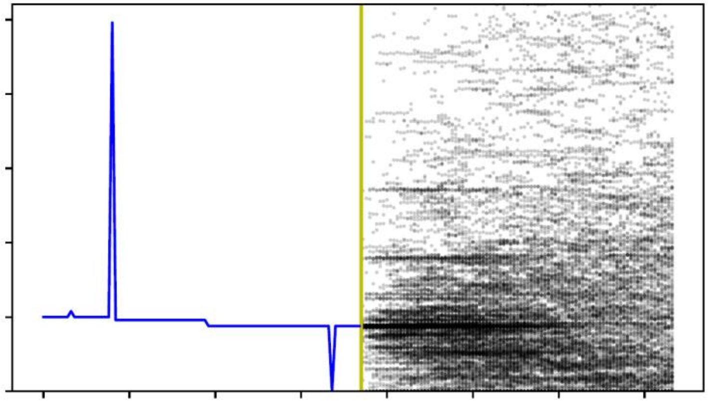
Time series have challenges not present in text/images



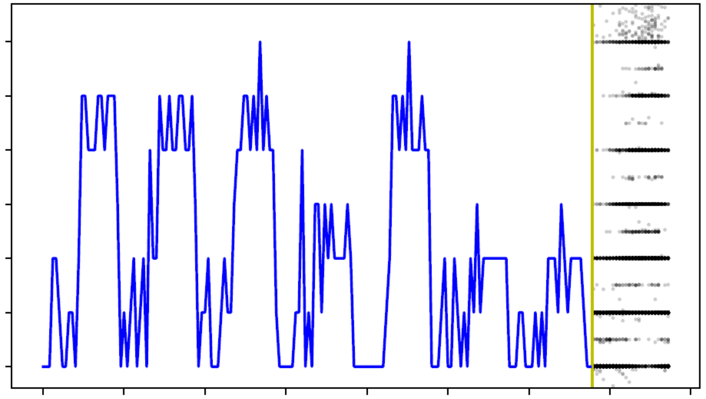
Trend, multiple seasonality



Arbitrary, multi-modal distributions



Unbounded/unknown dynamic range



Mix of continuous and discrete data

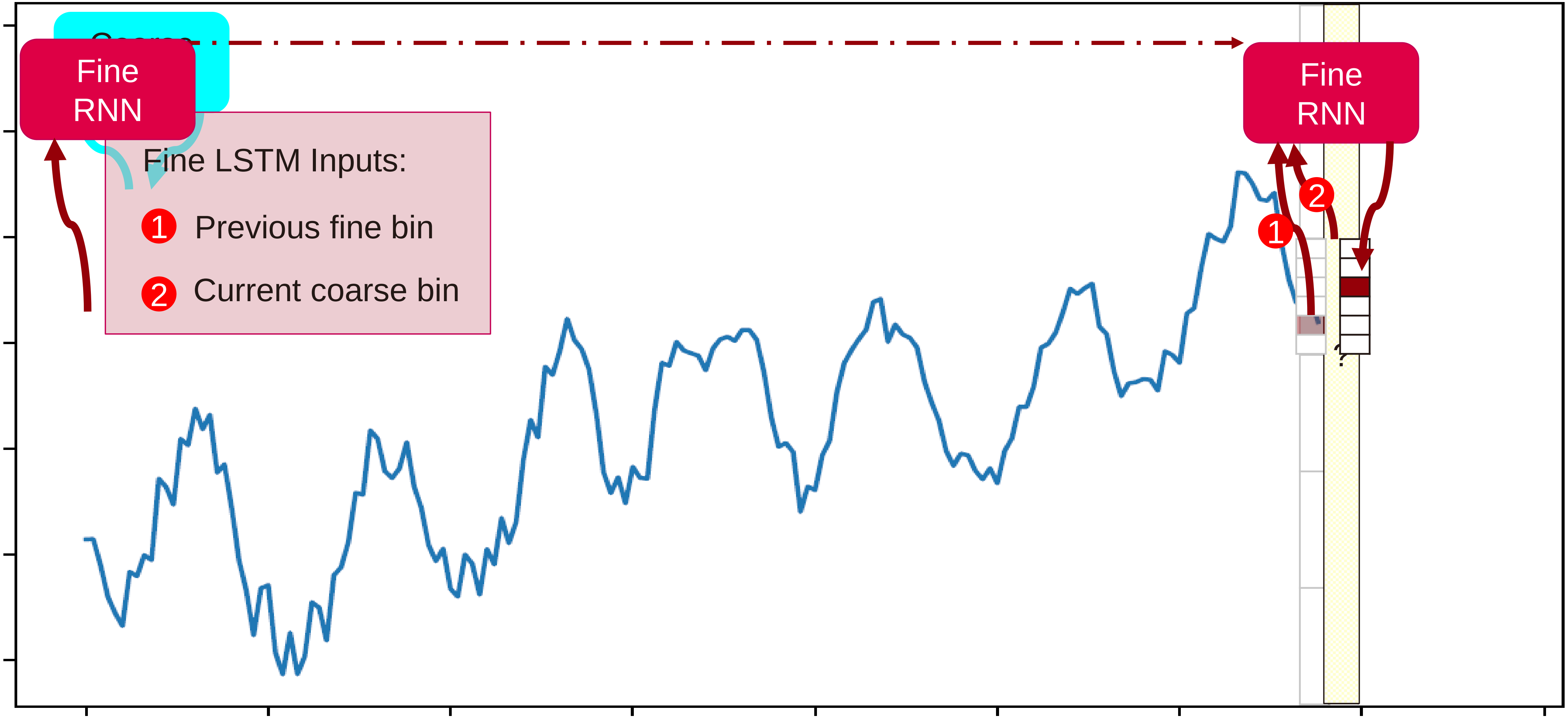
Practitioners advised to choose an output distribution appropriate for your dataset

- But different time series seem to require different output distributions

C2FAR solves this by generating precise distributions of arbitrary shape, without prior knowledge

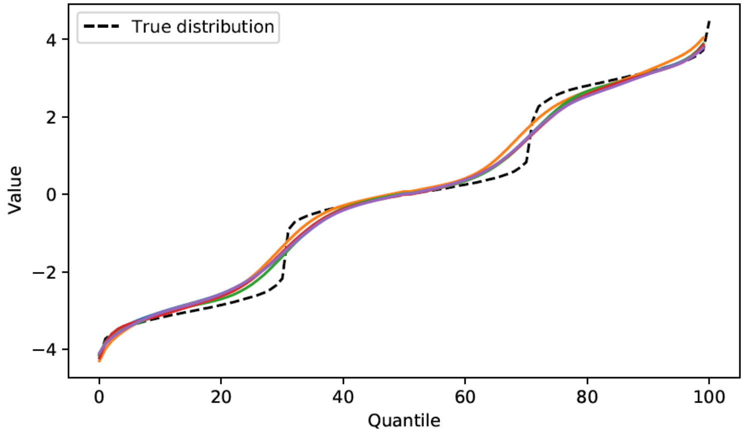
HUAWEI

C2FAR Probabilistic Forecasting (C2FAR-RNN)

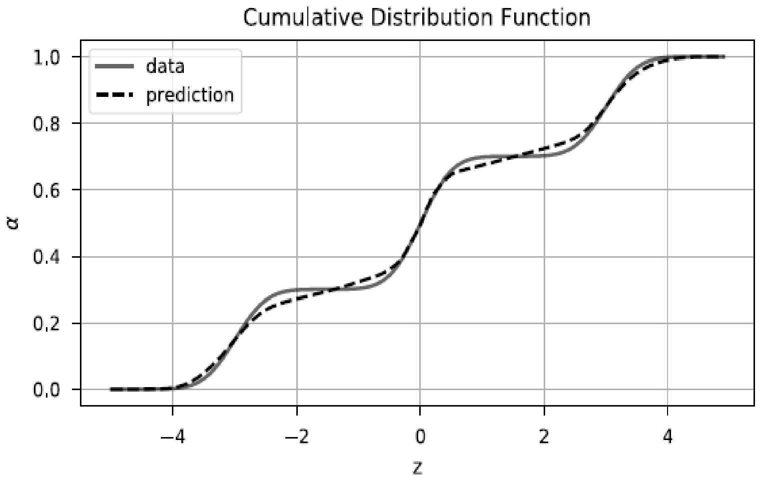


HUAWEI

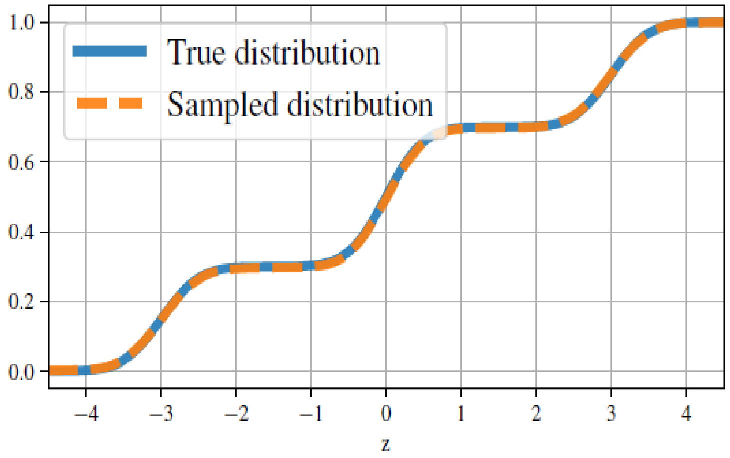
Synthetic Distribution Recovery



IQN-RNN (Gouttes et al., 2021)



SQF-RNN (Gasthaus et al., 2019)



C2FAR-RNN

✓ Unlike prior work, C2FAR-RNN has no problem fitting any given synthetic data distribution (continuous, discrete, etc.).

Forecasting Accuracy

Table 1: ND, wQL, Cov80, and Cov99 for our implementations (top), results from [49] (middle, denoted †) and [28] (bottom, denoted ‡), where available. In all cases, flat binned C2FAR-RNN$_1$ improves on DeepAR-Gaussian, while deeper C2FAR-RNN$_2$ likewise improves over C2FAR-RNN$_1$. Results are generally superior to prior state-of-the-art output distributions in RNN-based forecasting.

	ND%				wQL%				Cov80%	Cov99%
	elec	*traff*	*wiki*	*azure*	*elec*	*traff*	*wiki*	*azure*	*azure*	*azure*
Naive	40.8	73.6	35.7	3.49	40.8	73.6	35.7	3.49	-	-
Seasonal-naive	6.97	25.1	33.2	3.67	6.97	25.1	33.2	3.67	-	-
ETS	8.61	33.3	34.3	3.46	8.40	31.5	32.5	2.97	85.5/10.5	96.3/**20.3**
DeepAR-Gaussian	7.05	16.1	43.8	3.60	5.60	13.7	54.7	3.06	89.9/16.9	98.0/37.7
C2FAR-RNN$_1$	6.14	13.0	24.6	2.95	4.87	10.7	21.3	2.41	83.6/**8.3**	98.5/32.2
C2FAR-RNN$_2$	6.09	**12.9**	24.2	2.86	4.83	**10.6**	**21.0**	2.31	**79.0**/8.5	98.4/29.1
C2FAR-RNN$_3$	**6.00**	13.3	**24.1**	**2.77**	**4.76**	10.9	**21.0**	**2.27**	86.0/8.9	**98.6**/32.7
DeepAR-Binned†	8.21	23.2	94.6	-	6.47	18.8	84.7	-	-	-
DeepAR-StudentT†	6.95	14.6	26.9	-	5.71	12.2	23.8	-	-	-
IQN-RNN‡	7.40	16.8	**24.1**	-	-	-	-	-	-	-
SQF-RNN‡	9.70	18.6	32.8	-	-	-	-	-	-	-
DeepAR-StudentT‡	7.80	21.6	27.0	-	-	-	-	-	-	-

New dataset of cloud demand, released with paper

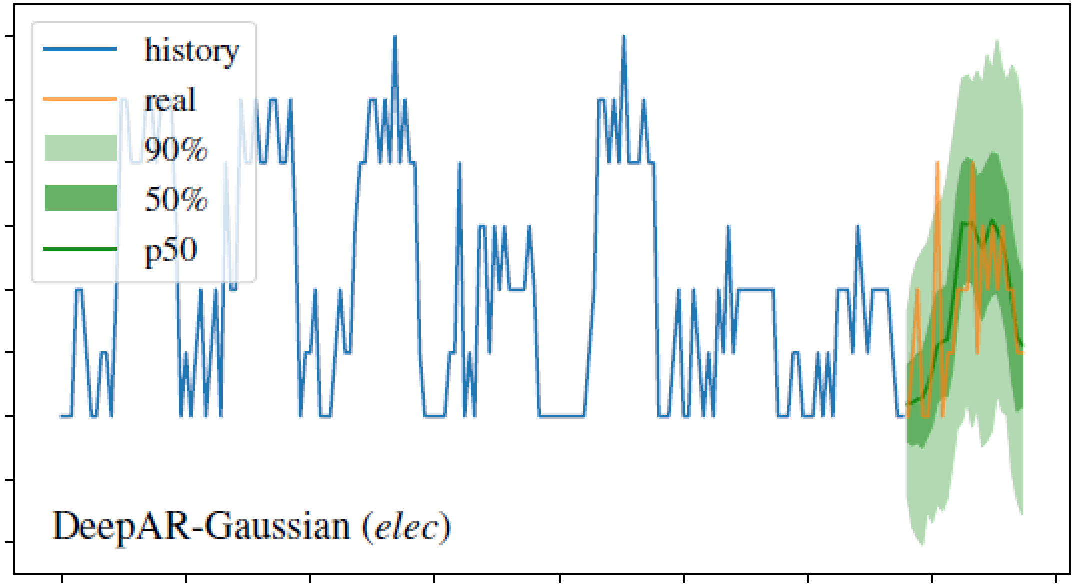
C2FAR:

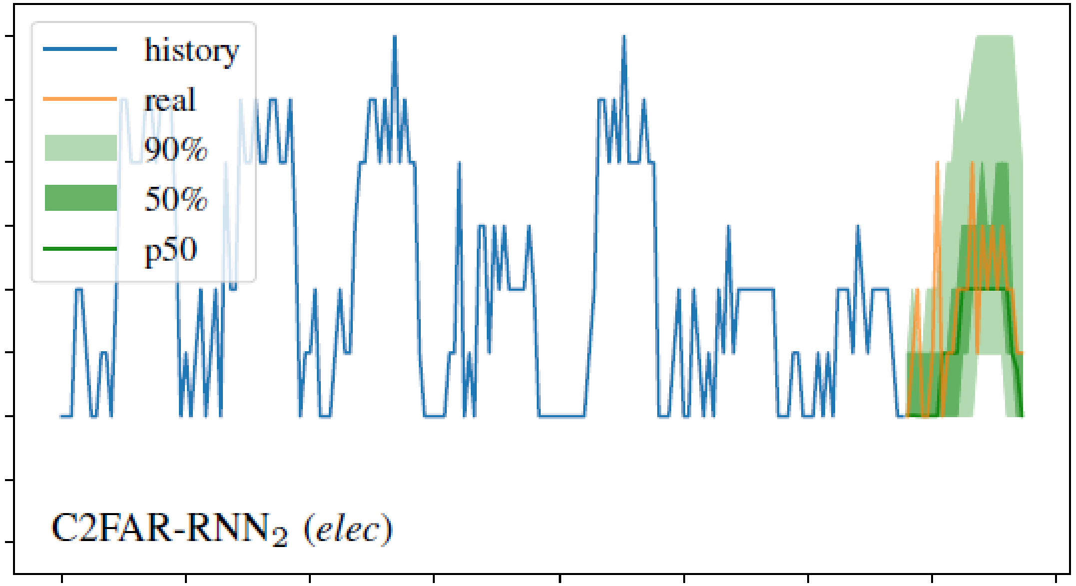
✓ Smaller absolute error

✓ Superior forecast quantiles

✓ Better-calibrated tails

Forecasting Quality





✓ C2FAR-RNN generates better forecast quantiles, suggesting higher-fidelity rollouts (better samples)

HUAWEI

Summary

C2FAR: turns generative modeling into a sequence of classifications over a hierarchical, discretized representation.

Code is available at https://github.com/huaweicloud/c2far_forecasting

HUAWEI