Estimating Interventional Outcomes over Time with Causal Normalizing Flow

Yoonseok Yeom[1*], Jonghwan Kim[1*], Taehui Yun[1], Juhyun Lyu[2], Jung-Hee Kim[2], Sangmin Lee[2], Jinseok Yang[2], Hyemin Jung[2], Woohyung Lim[2†], Sanghack Lee[1,3†]
([1]Seoul National University, [2]LG AI Research, [3]SNU-LG AI Research Center)

Seoul National University Graduate School of Data Science

uai2026

LG AI Research

Motivation

- **Causal inference over time** is central to healthcare and policy: *what happens under a hypothetical treatment strategy?*
- Most time-series causal methods return **point estimates** — they miss the variability that decision-making requires.
- Causal generative models estimate **interventional distributions**, but existing approaches are largely restricted to static settings.

Background

- **Structural causal models (SCMs)** characterize a data-generating process (DGP) through structural assignments with exogenous variables, inducing a causal graph $\mathcal{G}$:

$$X^j = f^j(\mathbf{Pa}^j, Z^j), \quad \text{for } j = 1, \dots, d$$

- **Causal normalizing flows (CNFs; Javaloy et al., 2023)** seek to recover an SCM by fitting an autoregressive flow aligned with $\mathcal{G}$.

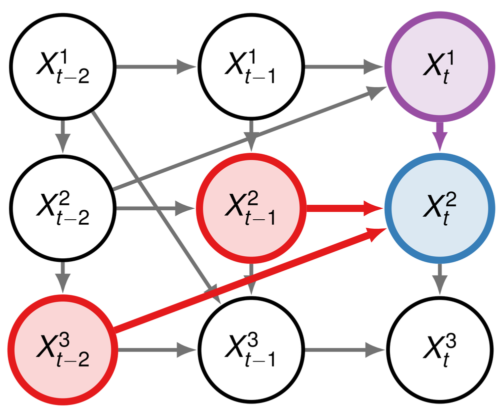
$$x^j = f_\theta^j(z^j;\ \mathbf{Pa}^j), \qquad \mathbf{z} \sim P(\mathbf{Z})$$

Problem Setup

For a multivariate time-series $\mathcal{X} \in \mathbb{R}^{T\times d}$, we assume its DGP admits a *Markovian*, *causally stationary* SCM.

Window causal graph $\mathcal{G}$ — induced by the SCM and assumed given:

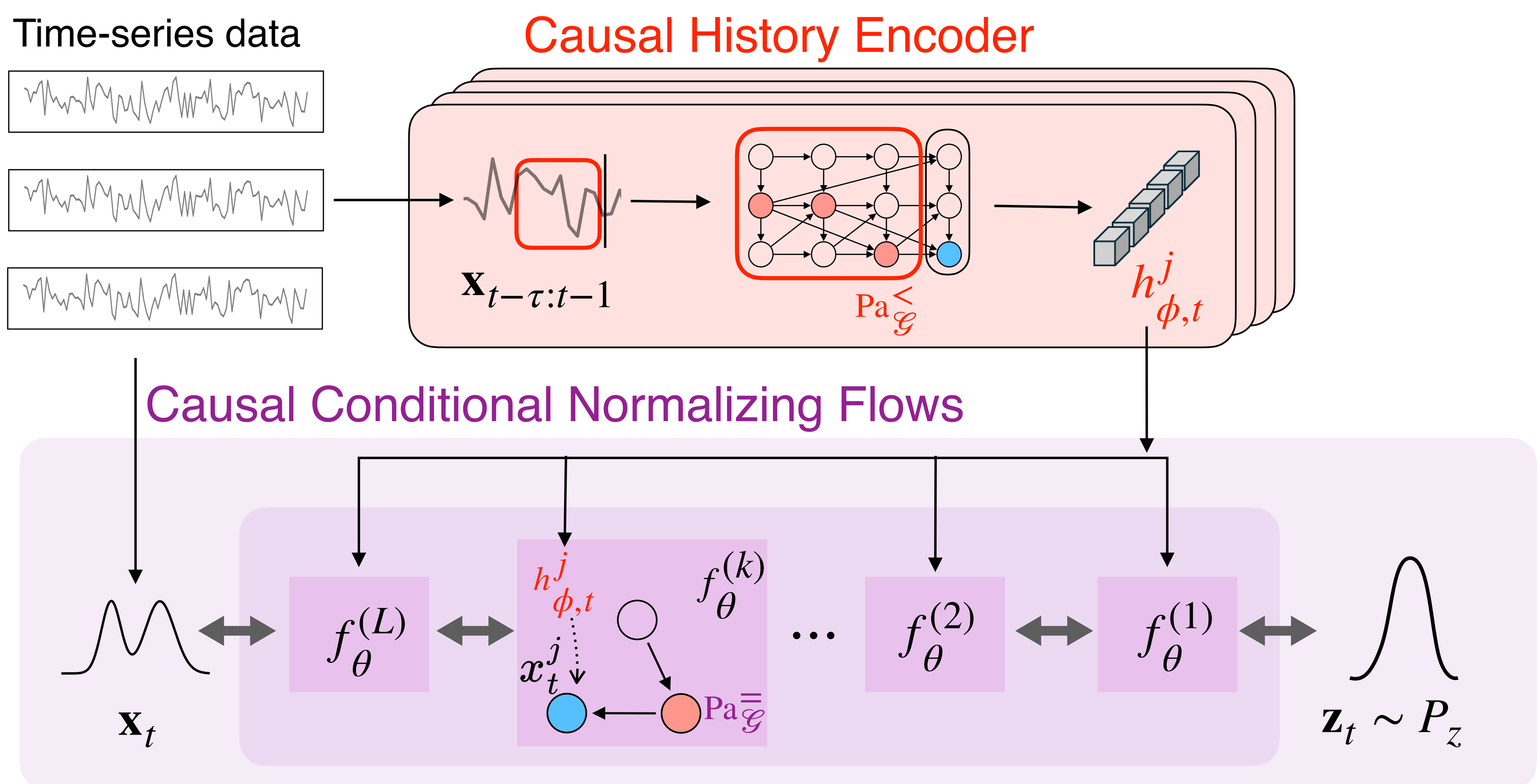
- $\mathrm{Pa}^{<}_{\mathcal{G}}(X_t^j)$ — time-lagged parents
- $\mathrm{Pa}^{=}_{\mathcal{G}}(X_t^j)$ — contemporaneous parents



Target estimand — the distribution of the future outcome Y_t under a hypothetical intervention sequence, conditioned on the history:

$$P(Y_t(\underline{a}_{m+1}) \mid \bar{\mathbf{x}}_m), \qquad t \geq m+1$$

TSCNF: Time-Series Causal Normalizing Flows



- **Causal History Encoder** — **encodes** only $\mathrm{Pa}^{<}_{\mathcal{G}}(X_t^j)$ into an embedding $h^j_{\phi,t}$ using feature-wise causal (parent) masking.
- **Causal Conditional Normalizing Flows** — **model** contemporaneous causal dependencies, with each X_t^j conditioned on its $h^j_{\phi,t}$.

$$x_t^j = f_\theta^j(z_t^j; \mathrm{Pa}^{=}_{\mathcal{G}}(X_t^j),\ h^j_{\phi,t}), \quad z_t^j \overset{\text{i.i.d.}}{\sim} P_z \quad \forall j$$

- **Trained** by maximizing the observational log-likelihood:

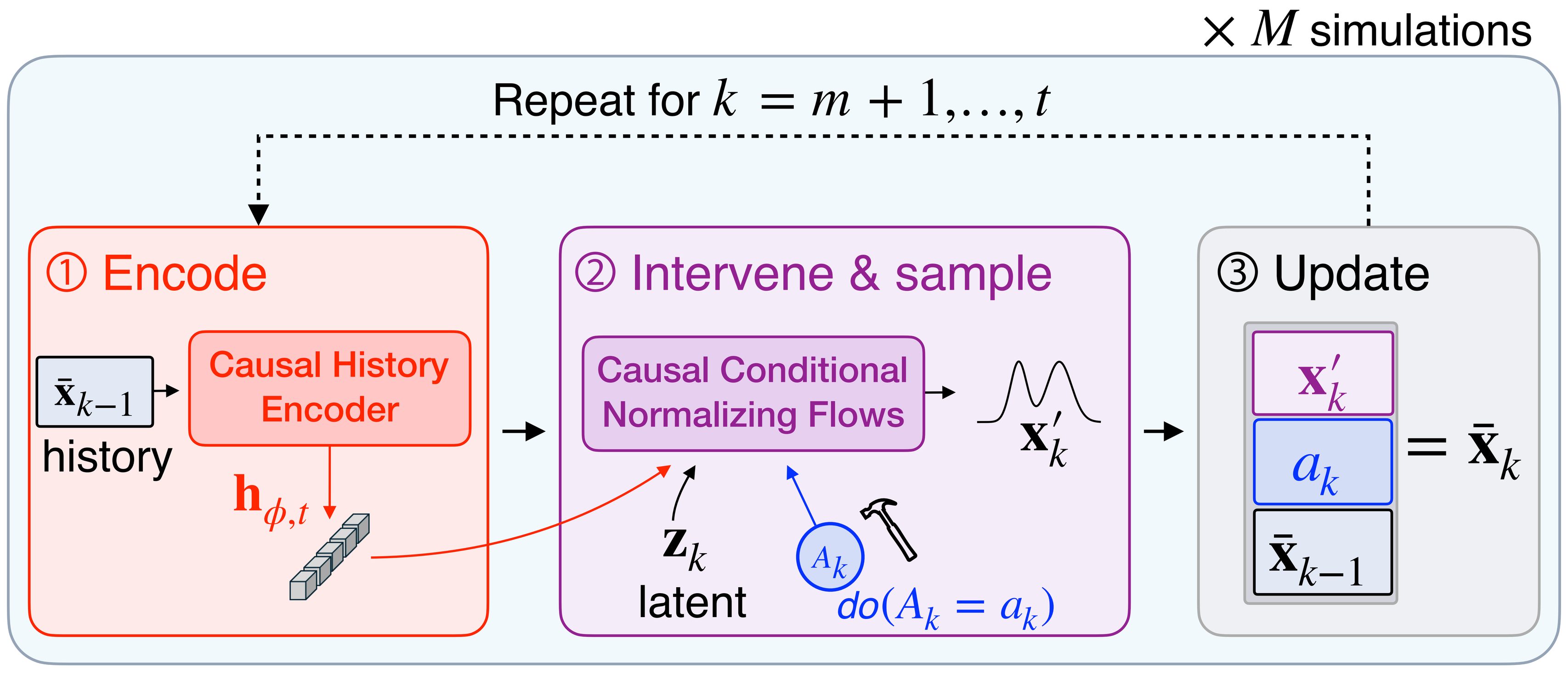
$$\max_{\theta,\phi} \sum_{t=1}^{T}\sum_{j=1}^{d} \log P_\theta\left(X_t^j \mid \mathrm{Pa}^{=}_{\mathcal{G}}(X_t^j),\ h^j_{\phi,t}\right)$$

Estimation of Interventional Distribution

- The target estimand admits the following form:

$$P(Y_t(\underline{a}_{m+1}) \mid \bar{\mathbf{x}}_m) = \int_{\mathbf{x}'_{m+1:t-1}} P(Y_t \mid do(a_t), \bar{\mathbf{x}}_{t-1}) \prod_{k=m+1}^{t-1} P(\mathbf{x}'_k \mid do(a_k), \bar{\mathbf{x}}_{k-1})$$

- Analytically intractable $\Rightarrow$ **Monte Carlo simulation** with TSCNF:



- At $k = t$, retain Y_t samples from $\mathbf{x}'_t$ to form the empirical distribution.

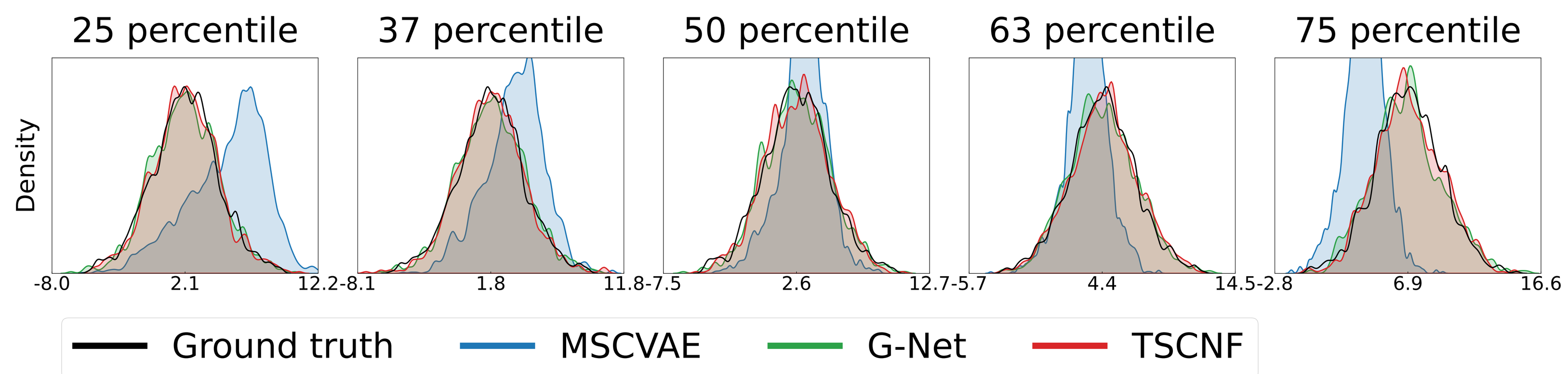
Experiments

A. Synthetic Dataset: Non-linear DGPs

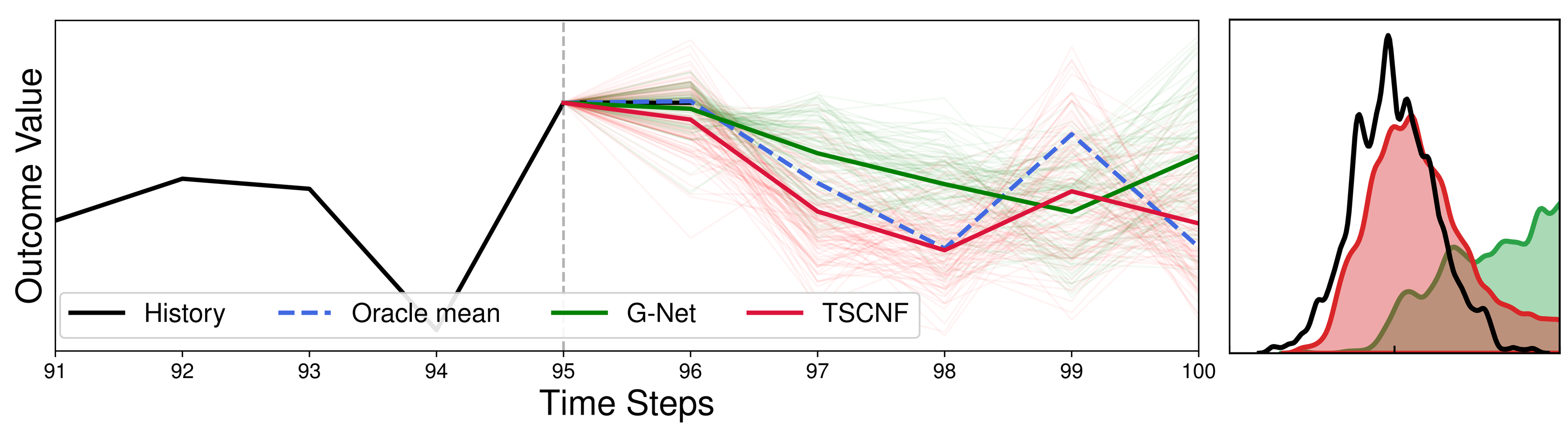
(a) Marginal query: TSCNF and G-Net outperform MSCVAE by avoiding unstable propensity weighting.

(b) History-adjusted query: Under sparse graphs, TSCNF outperforms G-Net by using only causally relevant signals.

(a) Marginal: $P(Y_{m+5}(\underline{a}_{m+1}))$



(b) History-adjusted: $P(Y_{m+5}(\underline{a}_{m+1}) \mid \bar{\mathbf{x}}_m)$



B. Tumor Growth Dataset: PK–PD Simulation (Geng et al., 2017)

- Point accuracy holds as the prediction horizon k grows.

Point estimation: $\mathbb{E}[Y_{m+k}(\underline{a}_{m+1}) \mid \bar{\mathbf{x}}_m]$

Model	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
RMSN	0.63 (0.11)	0.74 (0.12)	0.79 (0.11)	0.83 (0.11)	<u>0.88 (0.11)</u>
G-Net	**0.52 (0.03)**	<u>0.64 (0.04)</u>	<u>0.72 (0.06)</u>	<u>0.79 (0.09)</u>	0.92 (0.24)
Causal CPC	0.70 (0.10)	0.77 (0.10)	0.81 (0.10)	0.84 (0.12)	0.88 (0.14)
DoFlow	1.01 (0.05)	1.54 (0.06)	1.90 (0.06)	2.19 (0.07)	2.40 (0.09)
TSCNF	<u>0.55 (0.09)</u>	**0.58 (0.10)**	**0.63 (0.10)**	**0.67 (0.10)**	**0.71 (0.11)**

nRMSE % ; best in **bold**, second best <u>underlined</u>

Contributions

① **Extend causal normalizing flows** to the time-series setting.

② **Design a learning & inference framework** for multi-step interventional distribution estimation.

③ **Generalize across query types** without query-specific design.

Acknowledgements. This work was supported by LG AI Research and NRF (RS-2023-00211904) grant funded by the Korean government.