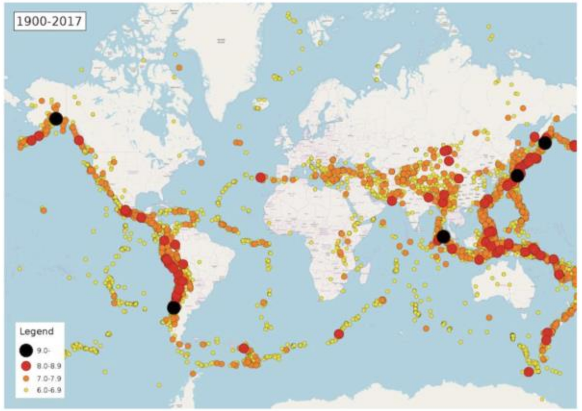
Spatio-temporal Diffusion Point Processes

苑苑
清华大学电子系城市科学与计算研究中心
2023/06/16

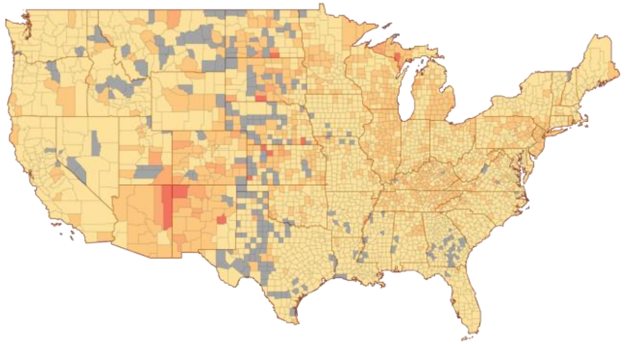
清華大學电子工程系
Department of Electronic Engineering, Tsinghua University

Spatio-temporal Events

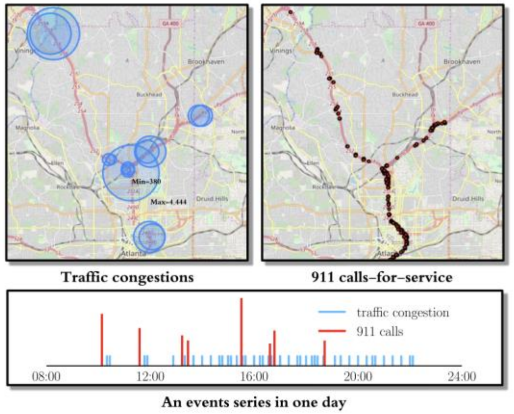
Events naturally come with time and location.



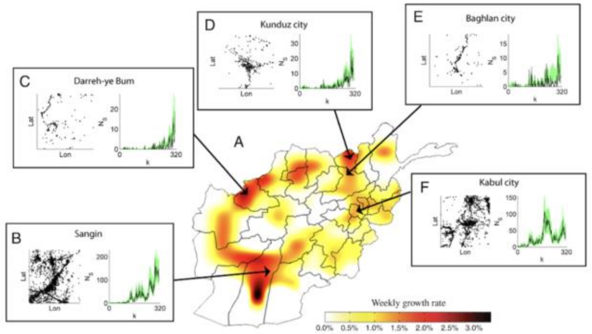
Earthquake



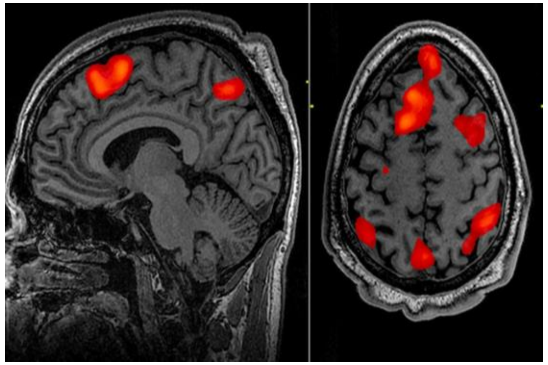
Covid-19



Traffic congestion



Conflict



Brain activity

Checkin

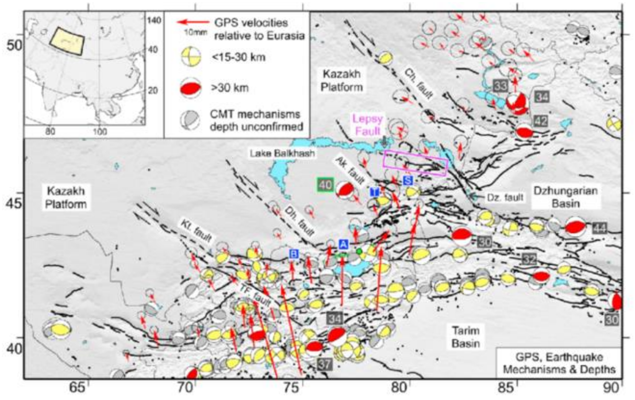
Background

Why we model spatio-temporal events?

时空事件随机性强，时空关联复杂度高，但其建模价值高、意义大

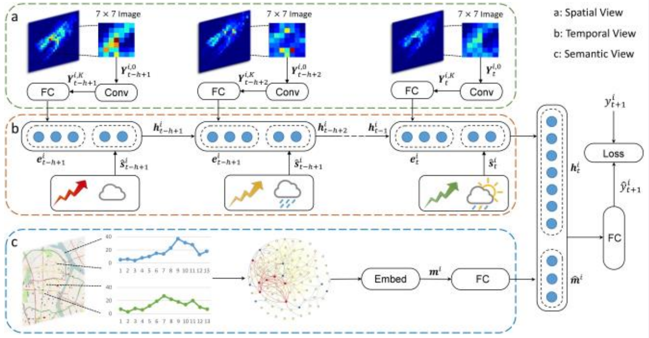
Understanding

Study the mechanisms that give rise to the dynamics of the recurrence of events



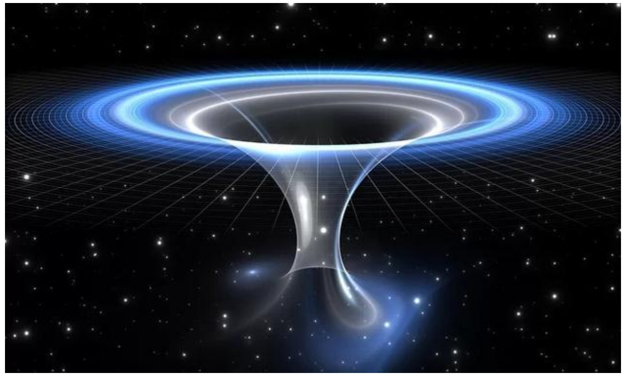
Predicting

Predict the dynamics of events in the future based on event history



Controlling

Design intervention and control measures to steer the dynamics of events to desirable outcomes



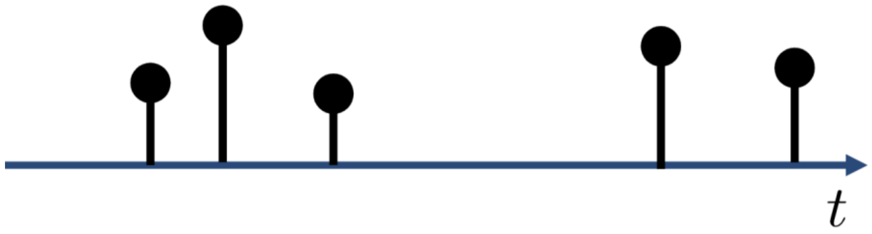
时空事件序列建模能有效预测未来事件，揭示其内在时空关联。

Background

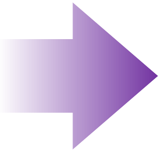
Event modeling: point processes

Discrete events localized in continuous time and space are usually modeled by point processes.

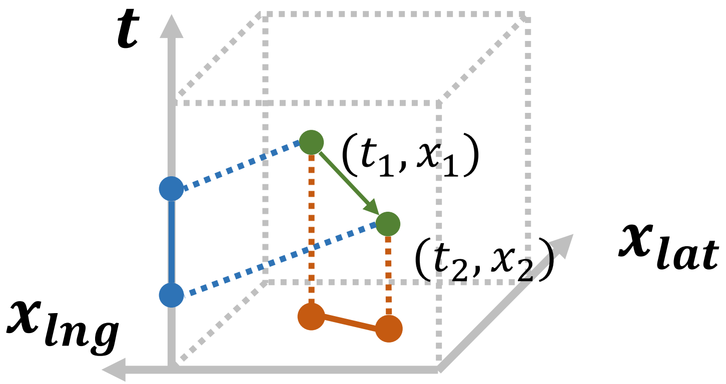
Point process in 1D
(Temporal Point Process)



A point process in time.



Point process in 3D
(Spatio-temporal Point Process)



A point process in space and time.

Background

Event modeling: temporal point processes (TPP)

- **Conditional Intensity function**

$$\lambda_k^*(t) = \lim_{\Delta t \downarrow 0} \frac{\Pr(\text{event of type } k \text{ in } [t, t+\Delta t) | \mathcal{H}_t)}{\Delta t},$$

- **Density**

$$f^*(t) = \lambda^*(t) e^{-\int_{t_{i-1}}^{t} \lambda^*(\tau) d\tau}$$

- **Likelihood**

$$\begin{aligned} L &= f^*(t_1) \cdot f^*(t_2) \cdot \cdots \cdot f^*(t_n) \cdot S^*(T) \\ &= \left(\prod_{i=1}^{n} \lambda^*(t_i) \cdot \exp\left(- \int_{t_{i-1}}^{t_i} \lambda^*(s) ds \right) \right) \cdot \exp\left(- \int_{t_n}^{T} \lambda^*(s) ds \right) \\ &= \left(\prod_{i=1}^{n} \lambda^*(t_i) \right) \cdot \exp\left(- \int_{0}^{T} \lambda^*(s) ds \right) \end{aligned}$$

From TPP to STPP

- **A STPP is uniquely defined by its conditional intensity $\lambda(t, x|H_t)$**
  - the rate of event occurrence given the history H_t

- **Likelihood**
  - $$\log p(\mathcal{H}) = \sum_{i=1}^{n} \log \lambda^*(t_i, \boldsymbol{x}_i) - \int_0^T \int_{\mathbb{R}^d} \lambda^*(\tau, \boldsymbol{x})\, d\boldsymbol{x} d\tau.$$

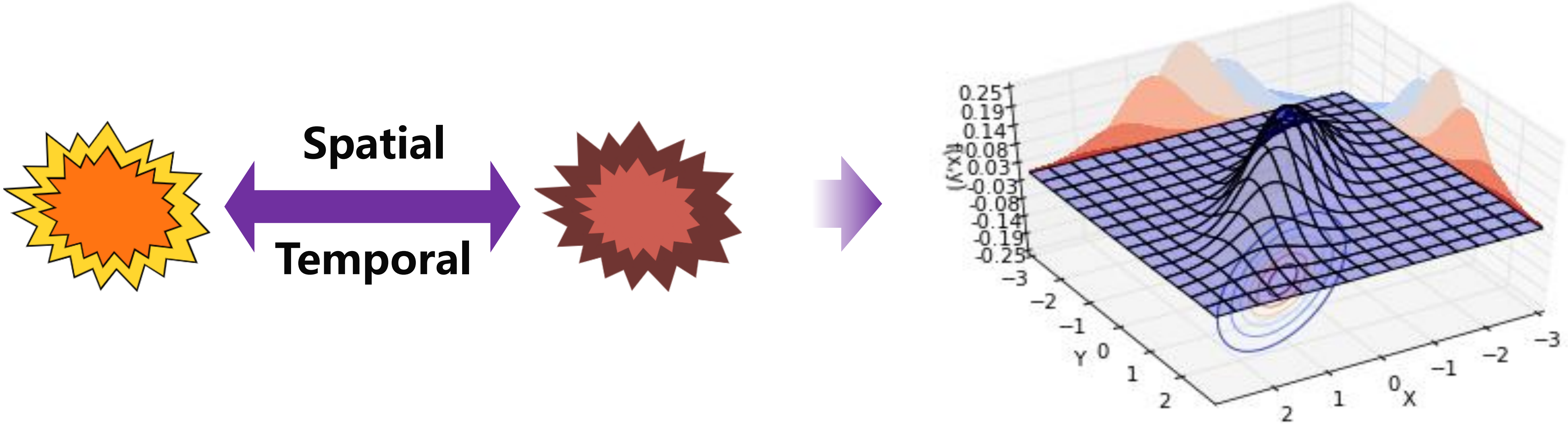
From TPP to STPP

Keep in mind when specifying the distribution f^*

- ✓ ***Closed-form likelihood*** : the likelihood should be computed in closed form for efficient model training.
- ✓ ***Closed-form sampling*** : Can we draw samples from f^* analytically?
- ✓ ***Flexibility*** : Does the given parametrization of f^* allow us to approximate any distribution?

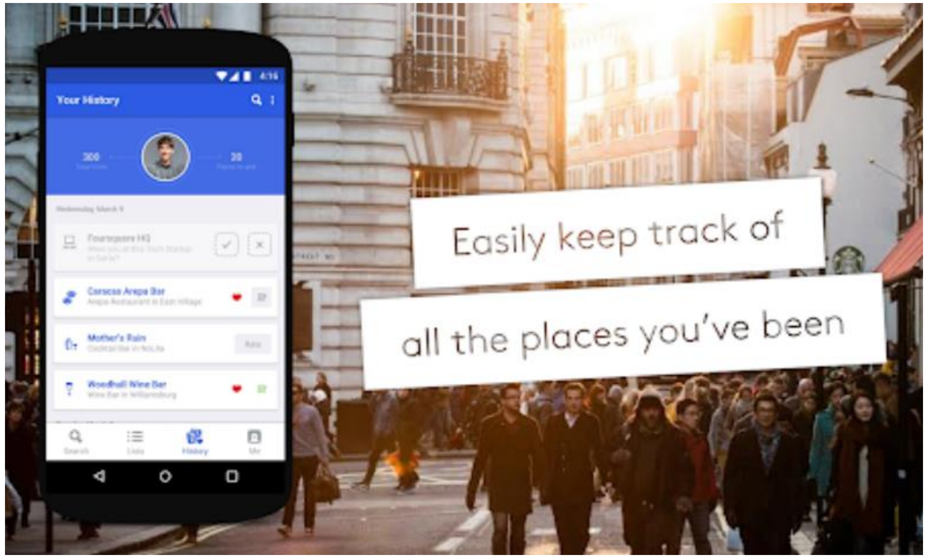
Challenges

Complex spatial and temporal correlations



Challenges

The spatial and temporal interaction patterns have high variations in urban environment.



Neural STPP

Neural Spatio-temporal Point Process, *NIPS 2020*

$$\lambda^*(t, \boldsymbol{x}) = \lambda^*(t)\, p^*(\boldsymbol{x} \mid t)$$

$$\log p(\mathcal{H}) = \sum_{i=1}^{n} \log \lambda^*(t_i, \boldsymbol{x}_i) - \int_0^T \int_{\mathbb{R}^d} \lambda^*(\tau, \boldsymbol{x})\, d\boldsymbol{x} d\tau.$$

Simplify

$$\log p(\mathcal{H}) = \underbrace{\sum_{i=1}^{n} \log \lambda^*(t_i) - \int_0^T \lambda^*(\tau)\, d\tau}_{\text{temporal log-likelihood}} + \underbrace{\sum_{i=1}^{n} \log p^*(\boldsymbol{x}_{t_i}^{(i)} | t_i)}_{\text{spatial log-likelihood}}$$

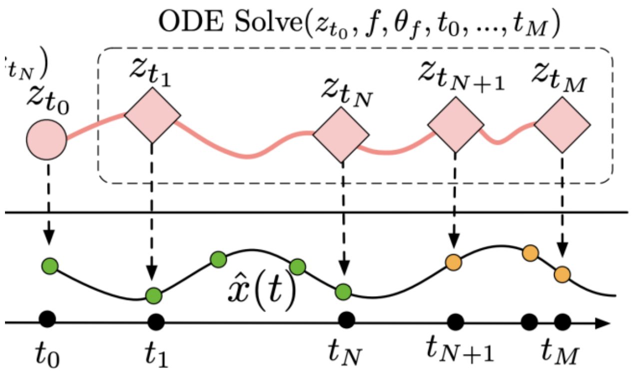
Neural STPP

Neural Spatio-temporal Point Process, *NIPS 2020*

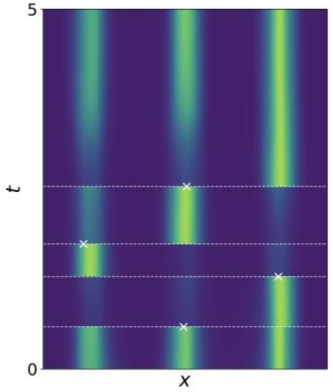
$$\lambda^*(t, \boldsymbol{x}) = \lambda^*(t)\, p^*(\boldsymbol{x} \mid t)$$

$$\boldsymbol{h}_{t_0} = \boldsymbol{h}_0$$
$$\frac{d\boldsymbol{h}_t}{dt} = f_h(t, \boldsymbol{h}_t)$$
$$\lim_{\varepsilon \to 0} \boldsymbol{h}_{t_i+\varepsilon} = g_h\left(t_i, \boldsymbol{h}_{t_i}, \boldsymbol{x}_{t_i}^{(i)}\right)$$

$$\boldsymbol{x}_0 \sim p(\boldsymbol{x}_0)$$
$$\frac{d\boldsymbol{x}_t}{dt} = f_x(t, \boldsymbol{x}_t, \boldsymbol{h}_t)$$
$$\lim_{\varepsilon \to 0} \boldsymbol{x}_{t_i+\varepsilon} = g_x(t_i, \boldsymbol{x}_{t_i}, \boldsymbol{h}_{t_i})$$



Neural ODE

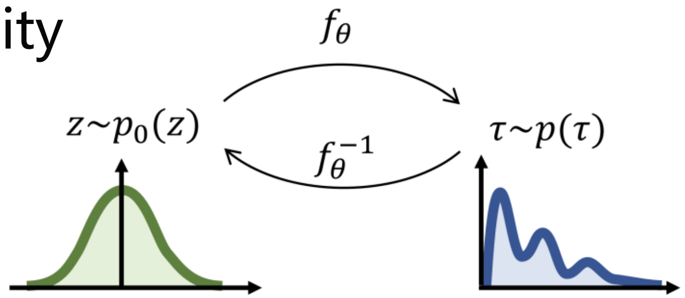


Continuous normalizing flow

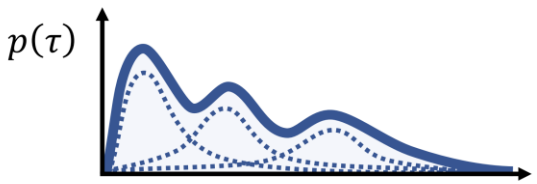
Limitation: require assumptions on the spatio-temporal dependence

Spatial distribution

- **Normalizing flows**
  - Get flexible distribution by transforming a simple density
  - $p(\tau) = p_0(f_\theta^{-1}(\tau)) \left|\frac{d}{d\tau} f_\theta^{-1}(\tau)\right|$
  - Invertible



- **Mixture distribution**
  - Convex combination of simple densities
  - $p(\tau) = \sum_k \pi_k p_0(\tau|\theta_k)$



They both require a certain structure in the functional approximators.

Trade off

Flexibility

Computation cost

Trade off

Flexibility — **Computation cost**

We expect the model:

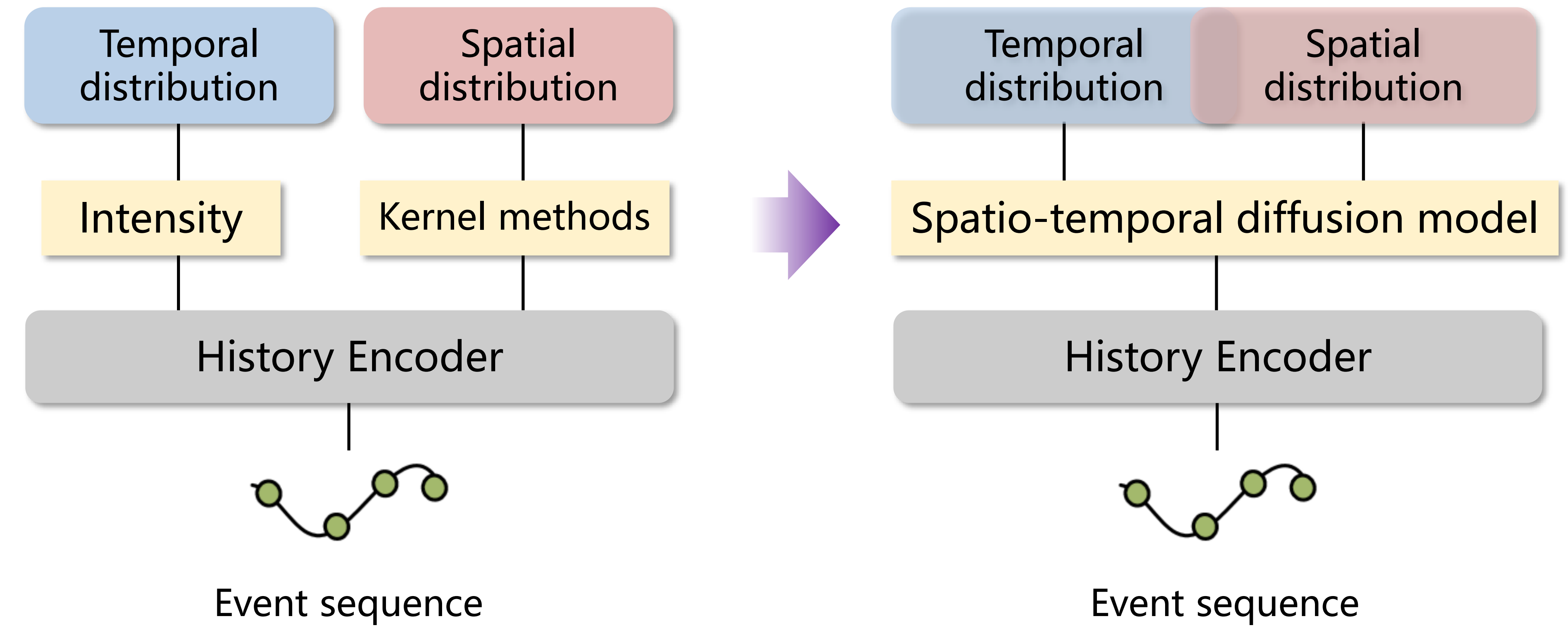
- ✓ **Update based on new events**
- ✓ **Easy to train and sample**
- ✓ **Use powerful neural networks freely**

Our Solution

Joint spatio-temporal distribution

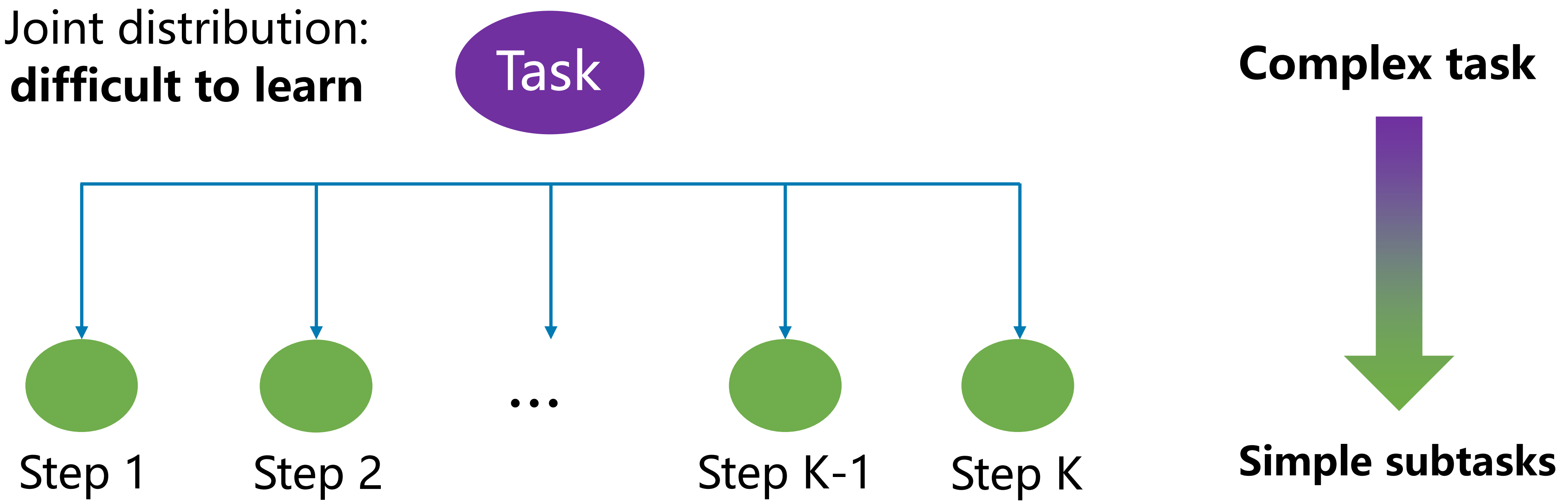
Conditionally independent models

Proposed model



Our Solution

From complex to simple



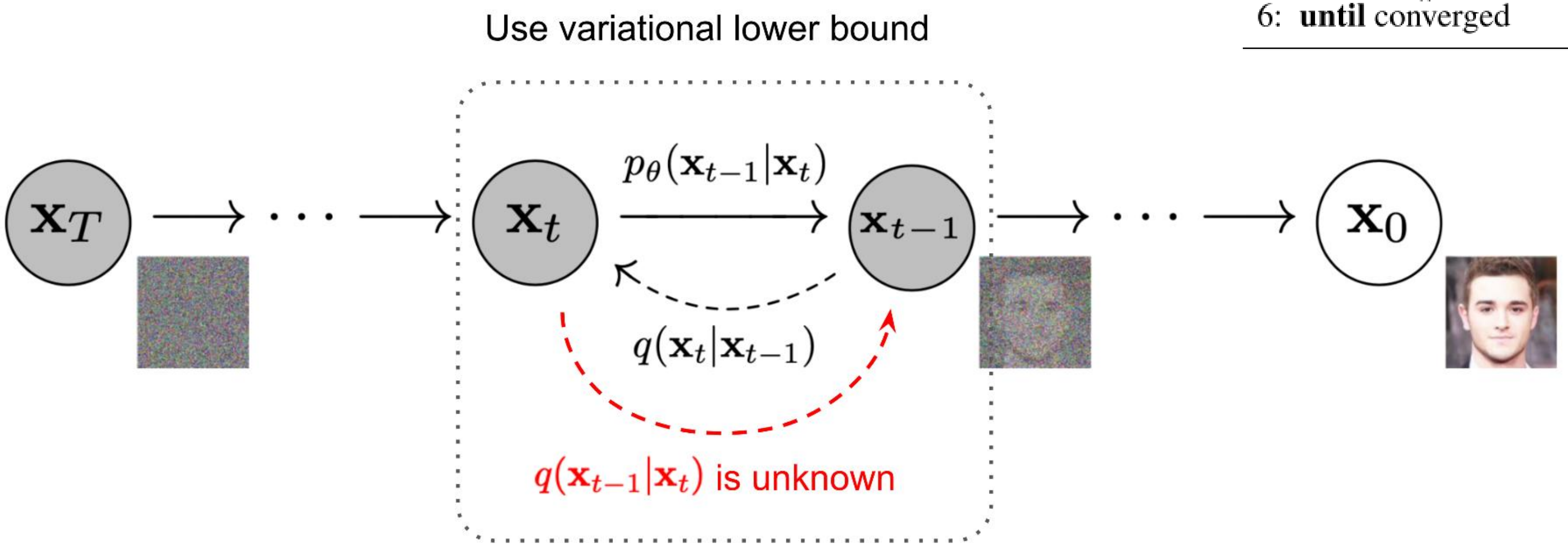
K subtasks: learn a **simple** distribution

Our Solution Diffusion Model

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar\alpha_t}\mathbf{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}, t) \right\|^2$
6: **until** converged

- **DDPM** [1]



Diffusion process: $q(\mathbf{x}_t|\mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$

Denoising process: $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$

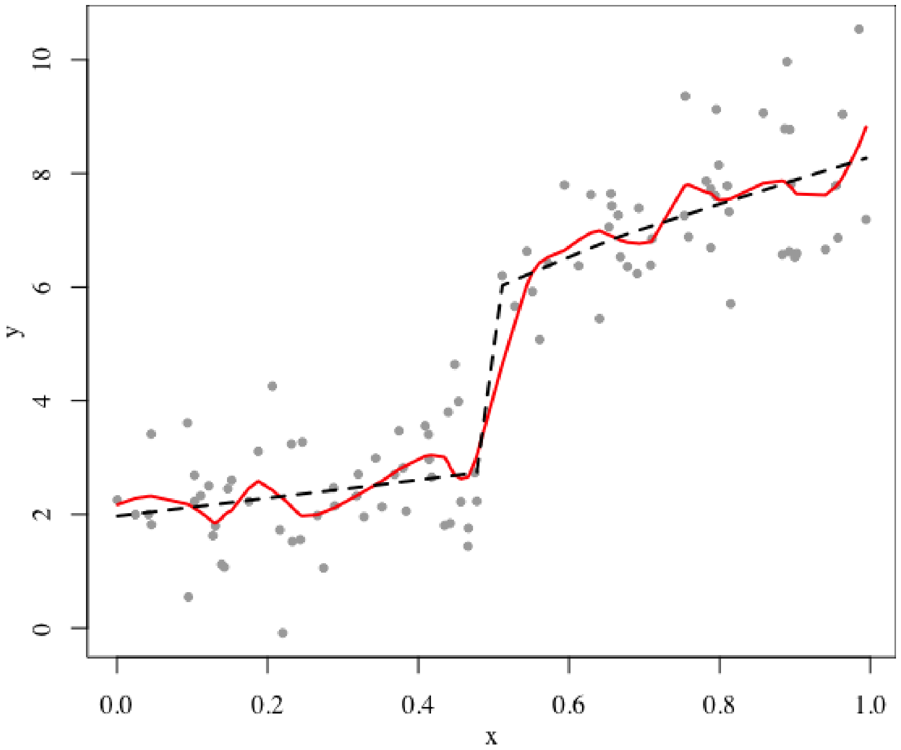
Training loss: $L_{\mathrm{simple}}(\theta) := \mathbb{E}_{t,\mathbf{x}_0,\boldsymbol{\epsilon}}\left[\left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar\alpha_t}\mathbf{x}_0 + \sqrt{1-\bar\alpha_t}\boldsymbol{\epsilon}, t) \right\|^2 \right]$

[1] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." NIPS 2020.

Our Solution

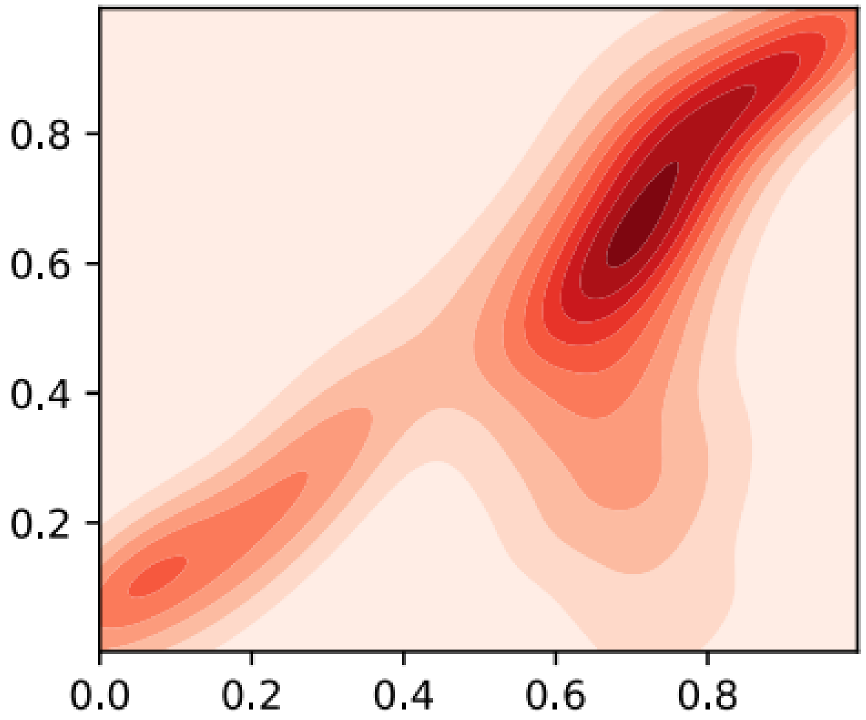
Why diffusion model is feasible?

Complex function



Similar idea

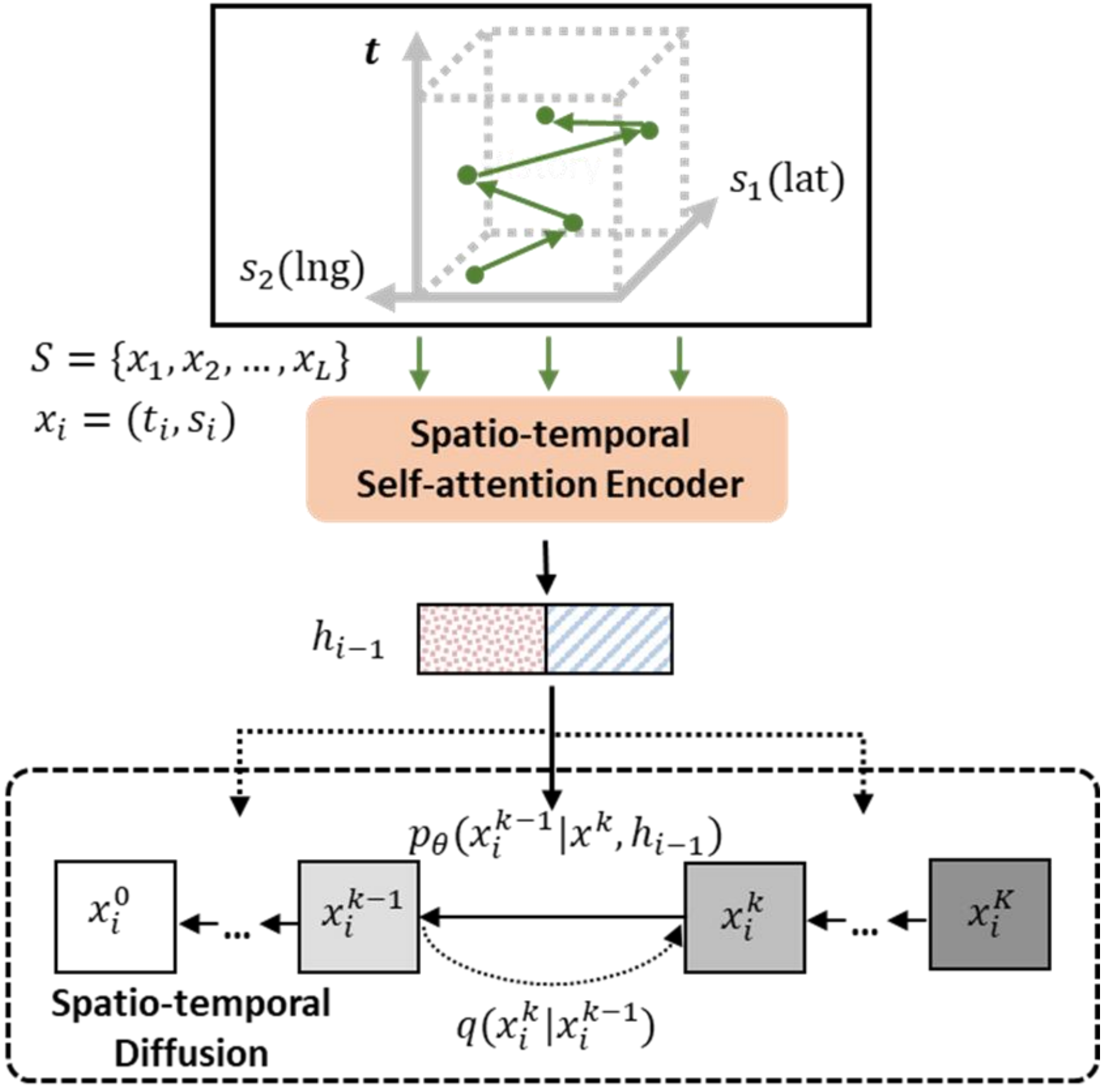
Complex distribution



Complex curves can be sufficiently fitted with piecewise linear function.

Minor changes can be sufficiently modeled by Gaussian distributions.

Our Solution Spatio-temporal Diffusion Point Processes



Spatio-temporal Self-att Encoder
Learn history representation

Condition

Spatio-temporal Diffusion Model
Learn ST joint distribution

Our Solution

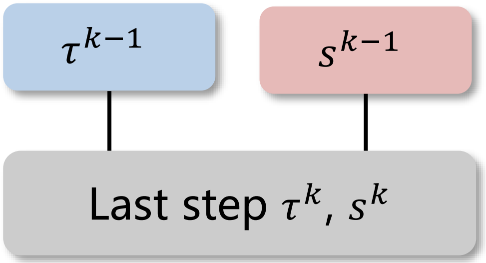
Spatio-temporal Diffusion Point Processes

- **Diffusion**
  - Separately diffuse on the spatial and temporal domains

$$q_{st}(\boldsymbol{x}_i^k|\boldsymbol{x}_i^{k-1}) := (q(\tau_i^k|\tau_i^{k-1}), q(s_i^k|s_i^{k-1})) \ ,$$

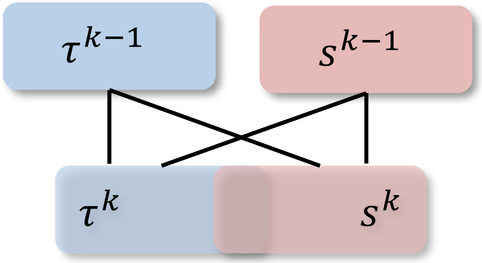
$$q(x^k|x^{k-1}) := \mathcal{N}(x^k; \sqrt{1-\beta_k}x^k, \beta_k \boldsymbol{I}) \ ,$$

Current step: independent



- **Denoise**
  - Conditionally independent at the current step
  - Dependent on the predicted values from the last step

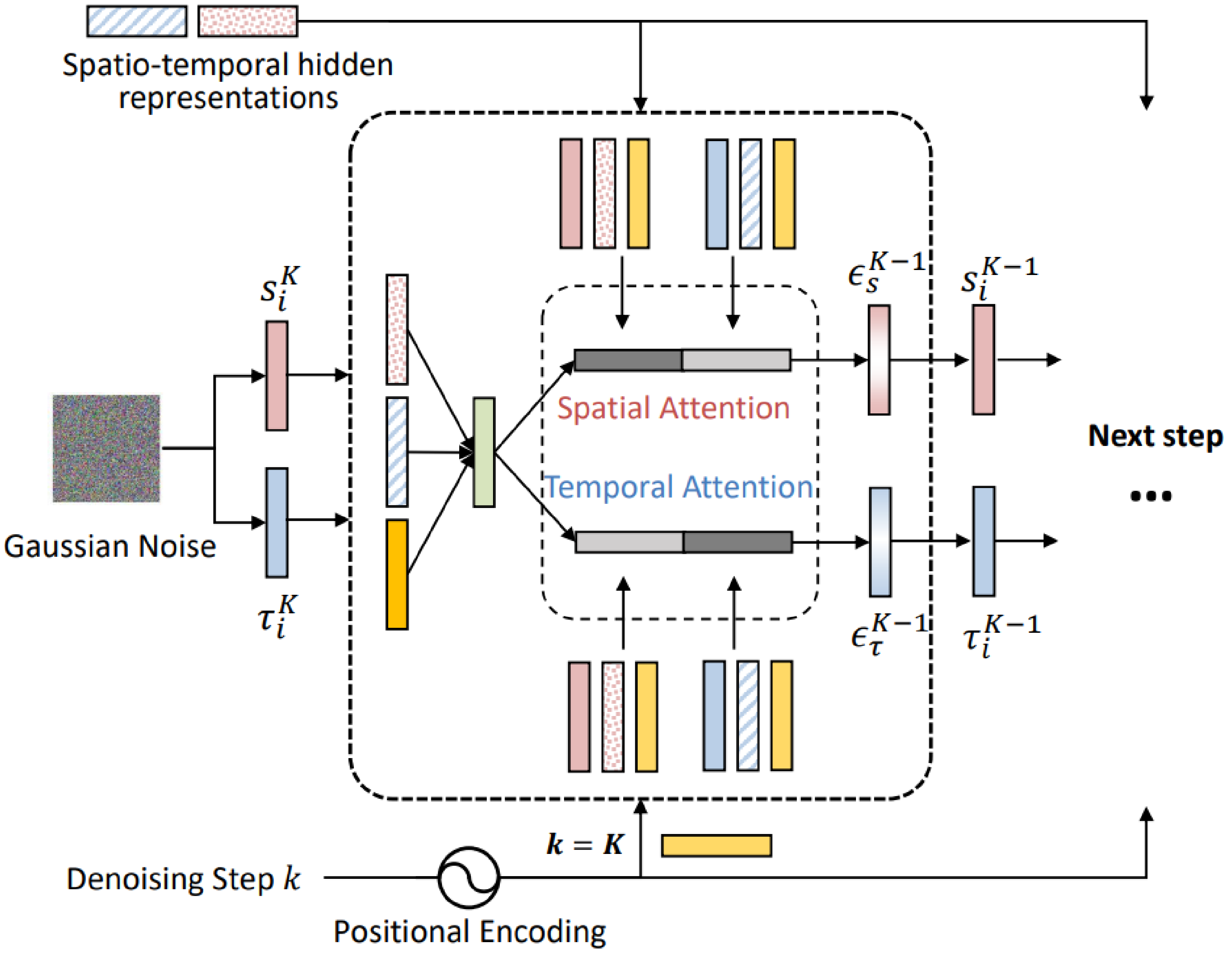
$$p_\theta(x_i^{k-1}|x_i^k, h_{i-1}) = p_\theta(\tau_i^{k-1}|\tau_i^k, s_i^k, h_{i-1})p_\theta(s_i^{k-1}|\tau_i^k, s_i^k, h_{i-1})$$



Interstep: interaction

Spatio-temporal Diffusion Point Processes

- **Parametrization of the denoising network**



Spatio-temporal Co-attention

Spatio-temporal Diffusion Point Processes

- **Comparison on important properties**

Model	No Asmp.(1)	No Restr.(2)	Flexible(3)	Closed-form sampling(4)
Hawkes [17]	✗	✗	✗	✗
Self-correcting [21]	✗	✗	✗	✗
KDE [2]	-	-	✗	✓
CNF [3]	-	-	✗	✓
ST Hawkes [45]	✗	✗	✗	✗
RMTPP [9]	✗	✗	✓	✗
NHP [32]	✗	✗	✓	✗
THP [69]	✗	✗	✓	✗
SNAP [64]	✗	✗	✓	✗
LogNormMix [47]	✗	✗	✗	✓
NJSDE [22]	✗	✗	✓	✗
Neural STPP [3]	✓	✗	✓	✗
DeepSTPP [66]	✗	✗	✓	✗
DSTPP (ours)	✓	✓	✓	✓

Datasets and Metrics

- **Datasets**
  - ***Continuous-space***
    - Real-world: Earthquake, Citybikes, Covid19, Human mobility
    - Synthetic: Pinewheel, HawkesGMM
  - ***Discrete-space***
    - Taxi-NYC
    - Crime-Atlanta
- **Evaluation metrics**
  - ***Log-likelihood***
  - ***Prediction error***
    - Temporal: MAE, RMSE
    - Spatial: Euclidean distance, Accuracy

Experiments

Overall performance

- **Likelihood**

Table 2: Performance evaluation for negative log-likelihood per event on test data. ↓ means lower is better. Bold denotes the best results and underline denotes the second-best results.

Model	Earthquake		COVID-19		Citibike		HawkesGMM	
	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓
Conditional KDE	2.21±0.105	-[(1)]	2.31±0.084	-	2.74±0.001	-	0.236±0.001	-
CNF	1.35±0.000	-	2.05±0.014	-	2.15±0.000	-	0.427±0.002	-
TVCNF	1.34±0.008	-	2.04±0.004	-	2.19±0.025	-	0.431±0.008	-
Possion	-	-0.146±0.000	-	-0.876±0.021	-	-0.626±0.000	-	1.34±0.000
Hawkes	-	-0.514±0.000	-	-2.06±0.000	-	-1.06±0.001	-	0.880±0.000
Self-correcting	-	13.8±0.533	-	7.13±0.062	-	7.11±0.010	-	4.59±0.135
RMTPP	-	0.0930±0.051	-	-1.30±0.022	-	1.24±0.001	-	1.52±0.002
NHP	-	-0.676±0.001	-	-2.30±0.001	-	-1.14±0.001	-	0.580±0.000
THP	-	-0.976±0.011	-	-2.12±0.002	-	-1.49±0.003	-	-0.402±0.001
SAHP	-	-0.229±0.007	-	-1.37±0.118	-	-1.02±0.067	-	-1.25±0.136
LogNormMix	-	-0.341±0.071	-	-2.01±0.025	-	-1.06±0.005	-	0.630±0.004
NJSDE	1.65±0.012	0.0950±0.203	2.21±0.005	-1.82±0.002	2.63±0.001	-0.804±0.059	0.395±0.001	1.77±0.030
NSTPP	0.885±0.037	-0.623±0.004	1.90±0.017	-2.25±0.002	2.38±0.053	-1.09±0.004	0.285±0.011	0.824±0.005
DeepSTPP	4.92±0.007	-0.174±0.001	0.361±0.01	-1.09±0.01	-4.94±0.016	-1.13±0.002	0.519±0.001	0.322±0.002
DSTPP (ours)	**0.308±0.006**	**-1.96±0.020**	**-1.02±0.029**	**-3.08±0.003**	**-5.41±0.011**	**-3.36±0.010**	**-2.95±0.047**	**-3.42±0.002**

[(1)] **Spatial baselines and temporal baselines can be combined freely for modeling spatio-temporal domains.**

Experiments

Overall performance

- **Event prediction**

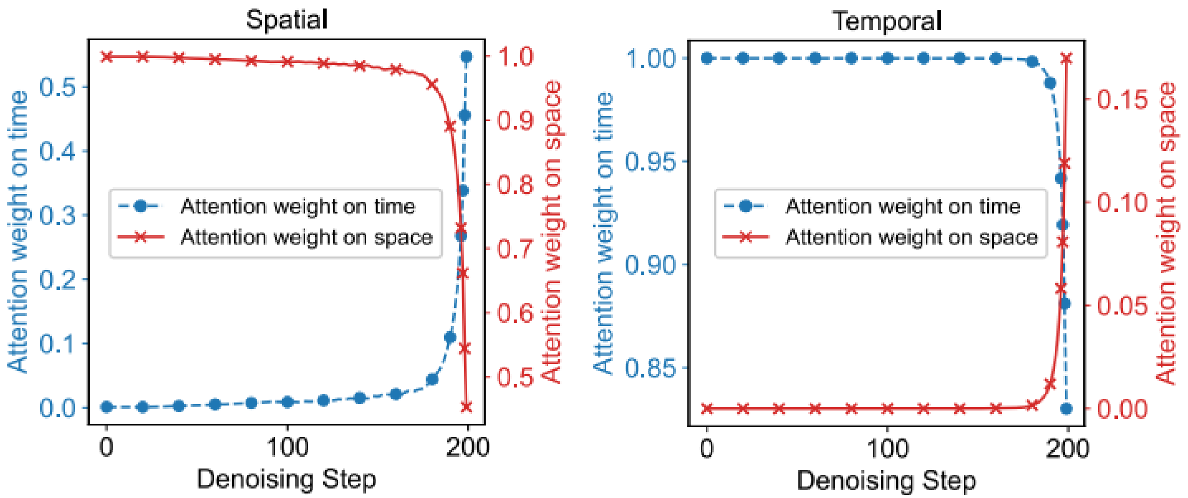
Table 3: Performance evaluation for predicting both time and space of the next event. We use Euclidean distance to measure the prediction error of the spatial domain and use RMSE between real intervals and predicted intervals for time prediction.

Model	Earthquake		COVID-19		Citibike		HawkesGMM	
	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓
Conditional KDE	11.3±0.658	-	0.688±0.047	-	0.718±0.001	-	1.54±0.006	-
CNF	8.48±0.054	-	0.559±0.000	-	0.722±0.000	-	71663±60516	-
TVCNF	8.11±0.001	-	0.560±0.000	-	0.705±0.000	-	2.03±0.000	-
Possion	-	0.631±0.017	-	0.463±0.021	-	0.438±0.001	-	2.81±0.070
Hawkes	-	0.544±0.010	-	0.672±0.088	-	0.534±0.011	-	2.63±0.002
Self-correcting	-	11.2±0.486	-	2.83±0.141	-	10.7±0.169	-	9.72±0.159
RMTPP	-	0.424±0.009	-	1.32±0.024	-	2.07±0.015	-	3.38±0.012
NHP	-	1.86±0.023	-	2.13±0.100	-	2.36±0.056	-	2.82±0.028
THP	-	2.44±0.021	-	0.611±0.008	-	1.46±0.009	-	5.35±0.002
SAHP	-	0.409±0.002	-	0.184±0.024	-	0.203±0.010	-	2.75±0.049
LogNormMix	-	0.593±0.005	-	0.168±0.011	-	0.350±0.013	-	2.79±0.021
WGAN	-	0.481±0.007	-	0.124±0.002	-	0.238±0.003	-	2.83±0.048
NJSDE	9.98±0.024	0.465±0.009	0.641±0.009	0.137±0.001	0.707±0.001	0.264±0.005	1.62±0.003	2.25±0.007
NSTPP	8.11±0.000	0.547±0.010	0.560±0.000	0.145±0.002	0.705±0.000	0.355±0.013	2.02±0.000	3.30±0.201
DeepSTPP	6.51±0.000	0.341±0.000	0.486±0.000	0.197±0.000	0.0312±0.000	0.234±0.000	1.38±0.000	1.46±0.000
DSTPP (ours)	**2.84±0.193**	**0.149±0.001**	**0.170±0.001**	**0.0243±0.000**	**0.00495±0.000**	**0.0301±0.002**	**0.136±0.013**	**0.0891±0.009**

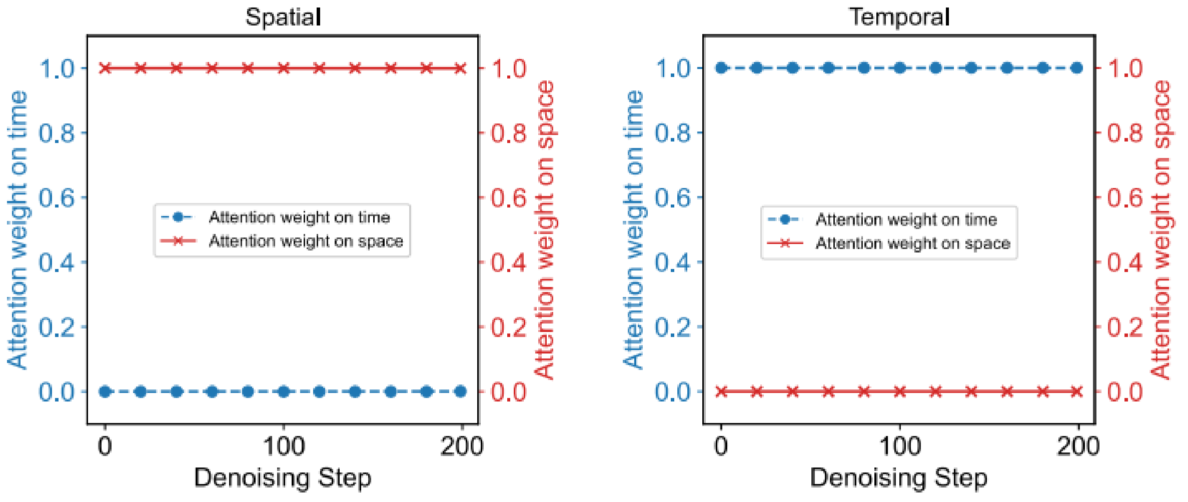
Experiments

Analysis of spatio-temporal interdependence

- **Comparison on important properties**



(a) Earthquakes



(b) Synthetic-Independent

***Spatio-temporal* dependent**

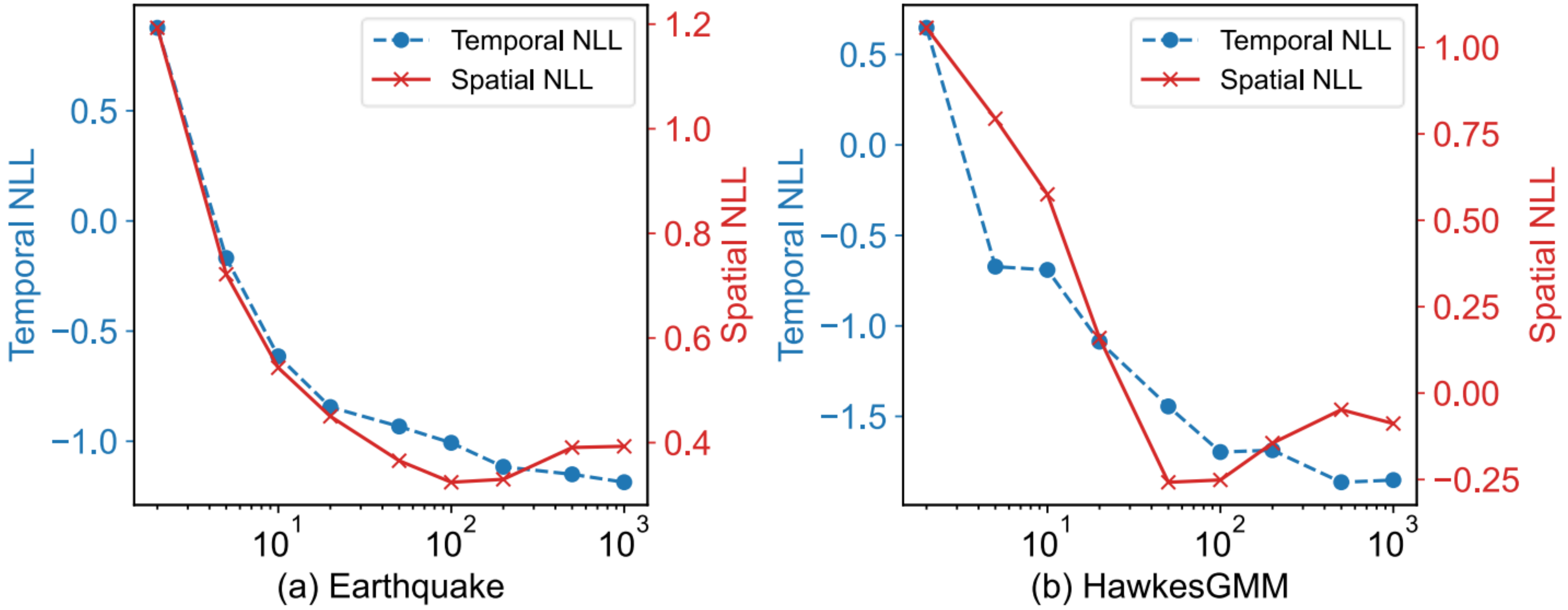
Different spatio-temporal relationships

***Spatio-temporal* independent**

Impact of diffusion steps

- **The number of total steps K is a crucial hyperparameter**

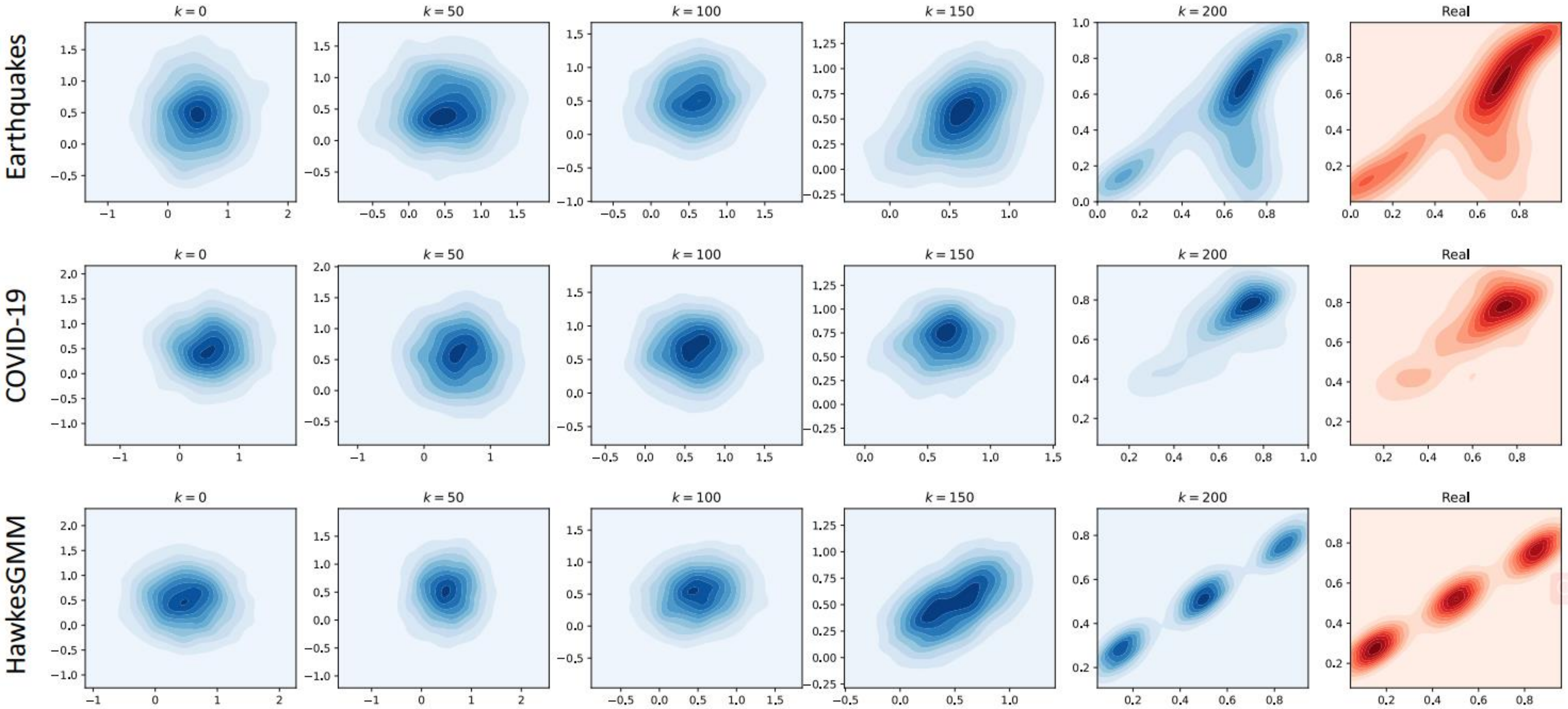
With the increase of diffusion steps, the denoising network approximates more minimal changes between steps.



Accuracy & Efficiency

Visualization of denoising processes

- **DSTPP is able to learn the generative process of distributions**



Thanks for listening !

https://github.com/tsinghua-fib-lab/Spatio-temporal-Diffusion-Point-Processes

Lab Info: http://fi.ee.tsinghua.edu.cn

Contact: liyong07@tsinghua.edu.cn