CAUSAL REINFORCEMENT LEARNING FOR SPATIO-TEMPORAL POINT PROCESSES

Anonymous authors

Paper under double-blind review

ABSTRACT

Spatio-temporal event sequences are increasingly accessible in various domains such as earthquake forecasting, crime prediction, and healthcare management. These data sources present unique challenges, as they involve both spatial and temporal dimensions, with event sequences exhibiting intricate dependencies over time and space. Neural network-based spatio-temporal point processes offer a sophisticated framework for modeling such event data. Conventional maximum likelihood estimation (MLE) of such data may lead to inaccurate predictions due to model-misspecification and compounding prediction errors. On the other hand, reinforcement learning frameworks, which treat event generation as actions and learn a policy to mimic event generation, may alleviate the training/test discrepancy issue. Current reinforcement learning of point processes may have prohibitively poor exploration efficiency. In this paper, we propose the Causal learning improved Reinforcement Learning Spatio-Temporal Point Process (CRLSTPP) framework, which can mitigate the issue of compounding prediction errors and improve exploration efficiency at the same time. Experiments on both synthetic data and real-world data validate the superiority of the proposed model.

025 026 027

004

010 011

012

013

014

015

016

017

018

019

021

1 INTRODUCTION

028 029

Asynchronous spatio-temporal event data, such as crime records, traffic flow patterns, and earthquake
occurrences, are widely utilized across various domains including predictive policing Zhu et al.
(2021b); Zhou et al. (2022); Mohler et al. (2011), urban transportation Yuan et al. (2023); Zhu
et al. (2021a), and seismology Chen et al. (2020); Ogata (1988). Different from regularly sampled
spatial or temporal data, the asynchronous sequence contains events with spatial coordinates and time
stamps that occur in continuous space and time. Those event sequences often demonstrate notable
self-exciting patterns. For instance, a crime event often leads to further incidents in neighboring
areas and time frames. It is critical to model these complex underlying dynamics such that accurate
predictions can be made sequentially given the history Baddeley et al. (2007); Chen et al. (2020);
Diggle (2006).

Spatio-temporal point processes (STPPs) are well suited for modeling event data of this nature. 040 Typically, an STPP model is characterized by an intensity function to indicate the instantaneous 041 rate of an event occurring at a specific location and time given all historical events. Classic Spatio-042 temporal point process models Moller & Waagepetersen (2003); Ogata (1988; 1998) leverage prior 043 process knowledge to make parametric or semi-parametric assumptions on the intensity function. 044 Although these models tend to perform well when the underlying parametric assumptions align with the actual processes, in scenarios where the genuine generative process of events remains obscure, these parametric assumptions can become excessively restrictive and do not represent the underlying 046 dynamics well. Alternatively, variations of point processes have been proposed to model intensity 047 functions by incorporating nonparametric forms Yuan et al. (2023), automatic integrationZhou & Yu 048 (2024), neural ordinary differential equations Chen et al. (2020), and recurrent neural networks Yang et al. (2018). 050

Conventionally, STPP models are trained via supervised learning, and model parameters are estimated
 by maximizing the log-likelihood (MLE) of observed event sequences. However, models learned via
 MLE at training time may be misspecified and may encounter compounding prediction errors over
 long horizons of composed forecasting Arora et al. (2022); Tan et al. (2018); Li et al. (2023), where

prediction errors made during testing accumulate in multiple successive forecasting steps and lead to
states distant from training data. Alternative methods address the training/test discrepancy problem
by employing reinforcement learning (RL) techniques Ranzato et al. (2015); Ding & Soricut (2017);
Bahdanau et al. (2016); Li et al. (2018). However, RL-based approaches may encounter issues with
the exploitation efficiency during the action sampling stage.

In this paper, we propose a novel causal reinforcement learning framework for modeling the spatiotemporal point process. Firstly, we introduce a new approach by examining spatio-temporal point processes from a reinforcement learning perspective. More specifically, the action is the next event's location and time, the state consists of the parameters that characterize the underlying spatial temporal distribution, and the generation of each event can be interpreted as an action taken by a stochastic policy. The observed sequences will be seen as the expert policy, and our goal is to use an MMDbased reward to minimize the discrepancy between the learner policy and the expert policy. Such a framework alleviates the training/testing discrepancy and reduces compounding prediction errors.

067 Secondly, to further enhance exploration efficiency, we measure the causal influence of actions 068 on the state and choose actions that are predicted to have a higher causal impact. Similar to the 069 self-exciting point process, the arrival action (event) will influence the underlying distribution. Our 070 goal is to measure this causal impact and to choose the action with the higher casual influence score. Additionally, unlike traditional discrete actions, our actions consist of event time stamps and spatial 071 locations, which are continuous. To tackle this issue, our method involves grid sampling to cover 072 the action space comprehensively; after that, we select the action that has the maximum deviation 073 from the expected outcome, enabling the policy to take the action that significantly improves the 074 exploration efficiency. Comprehensive experiments on both synthetic and real-world data demonstrate 075 that our proposed algorithm outperforms the state-of-the-art. 076

077

2 RELATED WORK

078 079

080 Spatial and Temporal Point Process Existing approaches in spatio-temporal point process modeling 081 such as kernel function-based models Zhang et al. (2023); Dong et al. (2023b); Zhu et al. (2021c); Ilhan & Kozat (2020); Dong et al. (2023a) often rely on simplified parametric assumptions to represent 083 the underlying dynamics of events such as seismic activities. Most of these solutions are based on the intensity paradigm, which might not be sufficiently flexible for different applications. Recently, some generative models Li et al. (2021b); Yuan et al. (2023); Xiao et al. (2017) have been proposed to 085 solve the training/testing discrepancy issue. However, the training complexity of generative models 086 is significant, which requires a substantial number of iterations and meticulous parameter tuning 087 to ensure that both neural networks achieve a state of equilibrium. Recent efforts aim to uncover 880 Granger causality in temporal point processes Jalaldoust et al. (2022); Xu et al. (2016). 089

A reinforcement framework Li et al. (2018) is proposed to model a temporal point process for 090 sequences with only temporal information. An imitation learning-based algorithm Zhu et al. (2021b) 091 has been proposed to train a spatio-temporal point process Zhu et al. (2021b). The work relies on fully 092 connected neural networks to parameterize a mixture of heterogeneous Gaussian diffusion kernels 093 for modeling spatio-temporal dependencies. These kernels capture complex spatial patterns, but the 094 model does not focus on capturing long-term temporal dependencies. Moreover, models based on 095 imitation learning also have some limitations due to their constrained exploratory capabilities Ali 096 (2021). In our work, we incorporate the recurrent mixture density network (RMDN) Bazzani et al. 097 (2016) structure, which effectively captures long-term dependencies between events.

098 **Causality Improved Reinforcement Learning** Incorporating causality information to enhance the performance of reinforcement learning (RL) frameworks is an emerging trend Zeng et al. (2023). 100 Recent studies have augmented RL performance by integrating causal learning in various ways. 101 Examples include utilizing the causal information to enhance sample efficiency in the online setting 102 Wang et al. (2021), using causality action inference as a part of reward bonus Seitzer et al. (2021), 103 and exploring actions with causal influence Lu et al. (2022); Seitzer et al. (2021); Bica et al. (2021). 104 Additionally, to improve the interpretability of the reinforcement learning model, some existing works 105 focus on learning the causal graph from the reinforcement learning process Ding et al. (2022); Li et al. (2021a). In this paper, we introduce causal information-based action exploration RL to the domain 106 of spatio-temporal point processes (STPP) for the first time, and we focus on causal information 107 detection to improve event prediction.

¹⁰⁸ 3 PRELIMINARIES

110 3.1 Spatio Temporal Point Process

112 A spatio-temporal point process is a stochastic process consisting of a sequence of events. The 113 i^{th} event can be represented as $a_i = (\mathbf{z}_i, t_i)$ where $t_i \in \mathbb{R}^+$ is the event occurrence time and 114 $\mathbf{z}_i \in \mathbb{R}^d$ is the event location in a *d*-dimensional space. Let \mathcal{H}_t denote the history up to time $t, \mathcal{H}_t =$ 115 $\{(\mathbf{z}_1, t_2), (\mathbf{z}_2, t_2), \dots, (\mathbf{z}_n, t_n) | t_n < t\}$. The spatio-temporal point process can be characterized by 116 the conditional intensity function:

$$\lambda(\mathbf{z}, t | \mathcal{H}_t) = \lim_{\Delta t, \Delta \mathbf{z} \downarrow 0} \frac{\mathbb{P}(t_i \in [t, t + \Delta t], \mathbf{z}_i \in B(\mathbf{z}, \Delta \mathbf{z}) | \mathcal{H}_t)}{|B(\mathbf{z}, \Delta \mathbf{z})| \Delta t},$$
(1)

where the $B(\mathbf{z}, \Delta \mathbf{z})$ denotes a Euclidean ball centered at $\mathbf{z} \in \mathbb{R}^d$ with radius $\Delta \mathbf{z}$, and the $|\cdot|$ is the Lebesgue measure.

The likelihood of observing an event at time t and location z is defined by:

$$\log p(\mathcal{H}) = \sum_{i=1}^{n} \log \lambda(\mathbf{z}_i, t_i | \mathcal{H}_{t_i}) - \int_0^T \int_{\mathbb{R}^d} \lambda(\mathbf{z}, t | \mathcal{H}_t) d\mathbf{z} dt.$$
(2)

The spatio-temporal point processes introduce additional complexity to the temporal ones since
 they require modeling the correlation between events across space. Training general STPPs with
 maximum likelihood poses a challenge due to the need to solve a complex multivariate integral.
 For computational tractability, approximations or numerical methods are often used to evaluate the
 integrals, especially over continuous space.

```
132
133
```

139

117 118 119

124 125 126

3.2 POLICY GRADIENT REINFORCEMENT LEARNING

Policy gradient is a category of reinforcement learning techniques where parameterized policies are learned to maximize the cumulative long-term reward through gradient descent. In contrast to value-based approaches, policy gradient methods focus on the direct optimization of the policy that guides the agent's actions. The objective in policy gradient methods is to maximize the expected return, defined as:

$$J(\theta) = \mathbb{E}_{\tau \sim \pi_{\theta}}[R(\tau)], \tag{3}$$

where θ represents the parameters of the policy π , τ denotes the trajectory, $R(\tau)$ is the total reward for trajectory τ . Additionally, the gradient of $J(\theta)$ with respect to θ is given by the policy gradient theorem:

$$\nabla_{\theta} J(\theta) = \mathbb{E}_{\tau \sim \pi_{\theta}} \left[\sum_{\tau=0}^{T} \nabla_{\theta} \log \pi_{\theta}(a_{\tau} | s_{\tau}) R(\tau) \right], \tag{4}$$

where a_t and s_t are the action and state at time step t, respectively, and T is the time horizon of the trajectory.

148 149

144 145

4 PROPOSED MODEL

150 151

152

4.1 REINFORCEMENT LEARNING OF STPP

In this paper, we focus on modeling intricate distributions of continuous time and space. We
employ the causal information-based reinforcement learning (RL) framework to model the STPP
by integrating the ideas from policy gradient reinforcement learning and casual influence detection
Seitzer et al. (2021).

157 We regard the observed spatio-temporal patterns as a series of actions enacted by the expert policy 158 π_E . Let $\xi = \{(\mathbf{z}_1, t_1), (\mathbf{z}_2, t_2), \dots, (\mathbf{z}_m, t_m)\}$ denote a spatio-temporal trajectory of *m* events from 159 the expert, where (\mathbf{z}_i, t_i) indicates an event at location \mathbf{z}_i and at time t_i . The event count *m* may 160 differ among various trajectories (sequences). Each event in the trajectory ξ can be seen as an action 161 sampled from the expert's spatio-temporal strategy π_E . Given a set of expert trajectories $E = \{\xi_1, \xi_2, \dots, \xi_k, \dots, \xi \sim \pi_E\}$, fitting a spatio-temporal point process to *E* can be seen as finding a



Figure 1: Our reinforcement learning framework is enhanced by causal learning. For each recurrent unit (excluding the initial one), the input consists of the output from the previous unit and the action generated by that previous unit. The observed sequences are treated as the expert policy, and the reward is used to minimize the discrepancy between the learner's policy and the expert policy.

learner policy π_L that is capable of generating a new set of trajectories $\widetilde{E} = \{\widetilde{\xi}_1, \widetilde{\xi}_2, \dots, \widetilde{\xi}_k, \dots | \widetilde{\xi}_k \sim \pi_L\}$ close to E.

Formally, consider a series of past events $\xi_t = \{(\mathbf{z}_1, t_1), \dots, (\mathbf{z}_i, t_i)\}_{t_i < t}$. Within this framework, the stochastic learner policy π_L samples a location and a time interval and generates the next event $(\mathbf{z}_{i+1}, t_{i+1})$. Following this sampling, the trajectory ξ_t updates to $\xi_{t+1} = \{(\mathbf{z}_1, t_1), \dots, (\mathbf{z}_i, t_i), (\mathbf{z}_{i+1}, t_{i+1})\}$ and the reward $r(\mathbf{z}_{i+1}, t_{i+1})$ will be calculated. Provided the reward function $r(\mathbf{z}, t)$, the optimal policy π_L^* can be deduced directly through:

182

183

185 186 187

188

189

198

199 200

where P is the set of all candidate spatio-temporal policies π_L , $\{(\mathbf{z}_1, t_1), \dots, (\mathbf{z}_{N^{\iota}}, t_{N^{\iota}})\}$ is a specific event sampled from policy π_L , and N^{ι} may differ for various samples.

 $\pi_L^* = \arg \max_{\pi_L \in P} J(\pi_L) := \mathbb{E}_{\iota \sim \pi_L} \left[\sum_{i=1}^{N^{\iota}} r(\mathbf{z}_i, t_i) \right],$

(5)

As shown in Figure 1, we adopt a recurrent mixture density network (RMDN)Bazzani et al. (2016) enhanced by stochastic neuronsHochreiter & Schmidhuber (1997) as the policy network to generate spatio-temporal events. To improve the efficiency of exploration, we adopt conditional mutual information (CMI) to measure the influence of potential actions and prioritize actions that are predicted to have a higher causal impact on the state.

206 207 208

4.2 POLICY NETWORK

The policy π_L needs to be sufficiently adaptable and comprehensive to encapsulate the intricate patterns of spatio-temporal point processes as observed in the expert's data. Thus, we utilize RMDN enhanced by stochastic neurons to accommodate the non-linear and extensive sequential dependencies inherent in the data.

214 Specifically, we assume the next time interval of an event follows Rayleigh distribution and the 215 location of an event follows the mixture of Gaussian distribution. The parameters of those distributions depend on the hidden state h_i at time t. Each recurrent unit corresponds to an event in the sequence, 216 and the unit is represented by the following functions: 217

$$f_i = \varphi(W_f \cdot [h_{i-1}, x_i] + b_f), \qquad p_i = \varphi(W_p \cdot [h_{i-1}, x_i] + b_p), \tag{6}$$

$$o_{i} = \varphi(W_{o} \cdot [h_{i-1}, x_{i}] + b_{o}), \qquad \qquad \widetilde{C}_{i} = \tanh(W_{C} \cdot [h_{i-1}, x_{i}] + b_{C}), \qquad (7)$$

228 229

233

239

243

247

250

251

218

$$C_i = f_i * C_{i-1} + p_i * \widetilde{C}_i, \qquad \qquad h_i = o_i * \tanh(C_i), \qquad (8)$$

222 where φ is the logistic sigmoid function and the f_t, p_t, o_t, C_t, h_t represent the forget gate, input gate, 223 output gate, memory cell, and hidden state, respectively. And the $x_i = \{z_i, t_i\}$ will be the action 224 generated by the previous unit.

225 To generate nonnegative-valued random time intervals, we adopt the Rayleigh distribution, and the 226 probability density function can be written as: 227

$$PDF(\psi(h_{i-1}),\alpha) = \frac{\psi(h_{i-1})}{\alpha^2} e^{-\frac{\psi(h_{i-1})^2}{2\alpha^2}},$$
(9)

230 where ψ is a nonlinear mapping from \mathbb{R}^d to the parameter space of the learner probability distribution, 231 and the α is the scale parameter of the Rayleigh distribution. With that, we can generate the time intervals as: 232

$$t_i \sim \pi(t|\psi(h_{i-1})). \tag{10}$$

234 After that, the mixture density network will take the hidden state h_i as the input and generate the parameters of the Gaussian mixture model, from which we sample a location \mathbf{z}_i . We can reparameterize the model as $\{(\mu_i^k, \phi_i^k, \sigma_i^k, \rho_i^k)\}_{k=1}^K$, where $\mu_i^k, \phi_i^k, \sigma_i^k$, and ρ_i^k are the 2D mean 235 236 position, the weight, the 2D variance and the correlation of the k-th Gaussian component, respectively. 237 The mixture model is therefore defined as follows: 238

$$y_i = \{ (\tilde{\mu}_i^k, \phi_i^k, \tilde{\sigma}_i^k, \tilde{\rho}_i^k) \}_{k=1}^K = W_y \cdot h_i + b_y,$$
(11)

240 where w_{y} and b_{y} are the weights to be estimated. We assume that the mixture Gaussian distribution 241 will decay with time, and the 2D location of the STPP can be sampled as: 242

$$\mathbf{z}_i \sim \pi(\mathbf{z}|y_i \cdot e^{-\beta(t_i - t_{i-1})}). \tag{12}$$

244 The stochastic action-generating component within our RMDN is designed to emulate the inherent 245 stochastic mechanisms of spatio-temporal point processes. The parameters of the spatial temporal 246 distributions such as μ, σ and ρ are the state in our reinforcement learning framework. Furthermore, once an action is generated, it will influence the underlying intensity distribution and have a cascading 248 effect on future events, similar to the dynamics observed in a self-exciting point process. 249

4.3 CAUSAL INFORMATION IMPROVED ACTION SELECTION

In this paper, we introduce the causal information based action selection method to enhance the exploration efficiency of the RL framework. Due to the self-exciting properties of spatio-temporal 253 point processes (STPP), in our RL framework, once an action is selected, it will influence the 254 subsequent state. Different actions exhibit varying degrees of influence on the state. To improve the 255 efficiency of exploration, our approach uses conditional mutual information to measure the influence 256 and prioritize actions expected to exert the most significant causal influence on the next state. 257

258 Firstly, unlike traditional discrete actions, our actions consist of event times and spatial locations, 259 which are continuous. We employ grid sampling to ensure comprehensive sampling while preserving the distribution's characteristics. We discretize the region of interest into $n \times n$ grid cells. Then we 260 generate actions from the learner policy π_L iteratively until q cells are covered by at least one event 261 sample. Finally, we extract q samples and each of them is located in a distinct grid cell. 262

263 Once an action $(a_i = (\mathbf{z}_i, t_i))$ is generated, the hidden states h and the parameters of RMDN 264 will change. Thus, the underlying spatio-temporal intensity will change. To encourage the action 265 exploration efficiency, we choose the possible event (action) that results in maximal conditional mutual information deviation from the expectation (see Appendix A for details). The process of 266 choosing the action can be described as: 267

$$a^* = \arg\max_{a \in \{a^1, \dots, a^z\}} \left(D_{KL} \left(p(y_i | y_{i-1}, a) \left\| \frac{1}{q} \sum_{k=1}^q p(y_i | y_{i-1}, a^q) \right) \right),$$
(13)

270 where the $a \in \{a^1, \ldots, a^q\} \sim \pi_L$, and y is the output of the RMDN as described in 11. Intuitively, 271 the selected action will be the one that causes the maximum deviation from the expected next state 272 among all possible actions. 273

274 4.4 REWARD FUNCTION 275

276 For spatio-temporal point processes, the complex patterns within the sequence cannot be adequately 277 scored using the L1 distance reward. Therefore, to effectively compare the STPP sequence, we adopt 278 an MMD reward Kim & Pineau (2013) function with a Gaussian kernel. The MMD measures the distance between two sequences, by mapping data points into a reproducing kernel Hilbert space 279 (RKHS) using a kernel function and computing the difference between their mean embeddings. The 280 definition is the following equation: 281

$$\mathsf{MMD}^{2}(\xi_{L},\xi_{E}) = \frac{1}{M_{L}^{2}} \sum_{i=1}^{M_{L}} \sum_{j=1}^{M_{L}} k(a_{i},a_{j}) + \frac{1}{M_{E}^{2}} \sum_{i=1}^{M_{E}} \sum_{j=1}^{M_{E}} k(e_{i},e_{j}) - \frac{2}{M_{L}M_{E}} \sum_{i=1}^{M_{L}} \sum_{j=1}^{M_{E}} k(a_{i},e_{j}),$$
(14)

where $a_i = (\mathbf{z}_i, t_i)$ is the learner policy action, e_i is the expert action, M_L is the event count of learner, and M_E is the event count of expert. Thus, our reward can be written as:

$$r(a) \propto -\sum_{u=1}^{N_E} \sum_{v=1}^{N_L} \text{MMD}^2(\xi_L^v, \xi_E^u),$$
 (15)

(16)

where N_E , N_L are the number of sequences generated by the expert and the learner, respectively.

4.5 POLICY GRADIENT

When utilizing Policy Gradient methods for training, the focus often lies on minimizing the divergence $D(\pi_E, \pi_L, \mathcal{H})$. However, it is equally viable to minimize $D(\pi_E, \pi_L, \mathcal{H})^2$ since the squaring function is a monotonic transformation. This approach falls within the reinforcement learning paradigm described in formulation 4. Employing the policy gradient technique, the gradient $\nabla_L D(\pi_E, \pi_L, \mathcal{H})^2$ is computed as:

300 301

282

284 285

292 293

295

296

297

298

299

302 303

304

305

306

 $\nabla_L J(L) \approx \frac{1}{M_E} \sum_{i=1}^{M_E} \left[\sum_{i=1}^{n_j} \left(\nabla_L \log \pi_L(a_i) \cdot r(a_i) \right) \right],$ where $\nabla_L \log \pi_L(a_i)$ denotes the gradient of the log-likelihood of an action a_i , drawn from a rollout sample $\xi = \{(\mathbf{z}_1, t_1), (\mathbf{z}_2, t_2), \dots, (\mathbf{z}_{N^{\iota}}, t_{N^{\iota}})\}$ under the learner's policy π_L . If we assume that spatial process and temporal process are independent of each other, we can simply multiply the two

307 Also, to reduce the variance of the gradient, we adopt the Mini-batch Reinforcement Learning Li 308 et al. (2018). For each training iteration, firstly, we sample a minibatch of M_l trajectories of events 309 from the expert policy π_E and a mini-batch of M_e trajectories from the learner policy π_L . These 310 trajectories are used to compute the policy gradient $\nabla_{\theta} D(\pi_E, \pi_L, \mathcal{H})^2$, which is an average over the 311 mini-batch gradients weighted by the estimated optimal policy π_L^* . The log-likelihood is obtained by 312 multiplying the spatial distribution and temporal distribution. Additionally, the policy parameters are 313 updated by moving in the gradient direction scaled by a learning rate η_l . This approach can diminish 314 the divergence between learner and expert policy distributions, thereby training the model to be close 315 to the trajectories in STPP.

316 317

318

EXPERIMENT 5

319 In this section, we conduct experiments to address the following questions: 320

- 321 **Q1:** How does the proposed model perform compared to existing baseline methods?
- **Q2:** What is the effect of causal learning on improving exploration efficiency? 322
- Q3: Has the model addressed the compounding prediction errors issue? 323

distributions to obtain the log-likelihood of action a_i .

Q4: How to attain a more profound comprehension of the RL-based STPP framework?

Table 1: Evaluation of performance in predicting both the time and location of the next 1 event.

	Syntl	hetic	Earth	quake	911	Call	Cri	me
Model	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial
CNF	-	0.102 ± 0.00	-	7.212±0.08	-	$0.218 {\pm} 0.01$	-	0.191±0.00
THP	$\underline{0.161{\pm}0.01}$	-	$0.412{\pm}0.03$	-	$0.230{\pm}0.04$	-	$0.601{\pm}0.04$	-
NEST	0.162 ± 0.01	0.099 ± 0.01	$0.415 {\pm 0.03}$	7.011 ± 0.61	$0.287 {\pm} 0.02$	$0.194{\pm}0.01$	$0.672 {\pm} 0.03$	0.193±0.00
NSTPP	$0.179{\pm}0.00$	$0.097 {\pm} 0.00$	$0.492{\scriptstyle\pm0.01}$	$6.983 {\pm} 0.11$	$0.232{\pm}0.01$	$0.173{\scriptstyle \pm 0.00}$	$0.611{\scriptstyle \pm 0.02}$	0.099±0.00
DeepSTPP	$0.198 {\pm} 0.01$	0.106 ± 0.00	$0.427{\scriptstyle\pm0.02}$	$6.747 {\pm} 0.53$	$0.202{\pm}0.01$	$0.188{\scriptstyle\pm0.00}$	$0.578{\scriptstyle\pm0.01}$	0.118 ± 0.00
DSTPP	$0.171{\scriptstyle\pm0.02}$	0.084±0.01	$\textbf{0.390}{\scriptstyle \pm 0.02}$	$\overline{6.915 \pm 0.75}$	$\overline{0.217{\pm}0.02}$	$0.178 {\pm} 0.02$	0.564±0.04	0.102 ± 0.02
CRLSTPP	$0.152{\pm}0.02$	0.089 ± 0.00	$\underline{0.393{\pm}0.02}$	6.608±0.58	$\textbf{0.197}{\scriptstyle \pm 0.01}$	$0.182{\pm}0.02$	0.581 ± 0.03	0.127 ± 0.01

Table 2: Evaluation of performance in predicting both the time and location of the next 20 events.

	Synthetic		Earthquake		911 Call		Crime	
Model	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial
CNF	-	$3.25{\pm}0.37$	-	302.53 ± 29.31	-	$6.03 {\pm} 0.77$	-	4.41±0.26
THP	$4.35{\scriptstyle\pm0.61}$	-	$11.13{\pm}2.06$	-	$8.014{\scriptstyle\pm1.83}$	-	$15.92{\pm}3.10$	-
NEST	$4.13{\scriptstyle\pm0.47}$	$2.55{\scriptstyle\pm0.35}$	9.44 ± 1.22	$231.25 {\pm} 28.34$	7.42 ± 1.10	$4.70{\pm}0.48$	$\underline{15.59}{\scriptstyle\pm2.36}$	3.98 ± 0.97
NSTPP	$4.60{\pm}0.31$	$3.71{\pm}0.30$	$11.17{\pm}1.24$	$274.14{\scriptstyle\pm25.53}$	$8.49{\pm}1.02$	$5.51{\pm}0.81$	$17.24{\scriptstyle\pm2.14}$	$4.71{\scriptstyle\pm0.61}$
DeepSTPP	$4.22{\pm}0.51$	$3.51{\pm}0.42$	$11.23{\pm}1.01$	$268.09{\pm}31.04$	$8.51{\scriptstyle\pm1.07}$	$5.72{\pm}0.92$	$16.71{\scriptstyle\pm2.46}$	$4.55{\scriptstyle\pm0.91}$
DSTPP	$4.28{\scriptstyle\pm0.72}$	$3.42{\pm}0.60$	$10.59{\scriptstyle\pm1.46}$	228.12 ± 33.47	$8.77{\pm}1.38$	$5.38{\scriptstyle\pm1.19}$	$15.98{\scriptstyle\pm3.06}$	$3.79{\pm}0.89$
CRLSTPP	$\textbf{3.29}{\scriptstyle \pm 0.84}$	$\pmb{2.02}{\pm}0.34$	$\pmb{8.02}{\pm}1.02$	201.28 ±22.19	7.02 ±1.14	$3.71{\pm}0.73$	$13.04{\pm}1.82$	3.28±0.83

351

361 362

363

324

326 327 328

5.1 DATASET

We utilize one synthetic data and three real-world datasets exhibiting diverse temporal dynamics across various domains, and further details can be found in Appendix B. Our code will be made available on GitHub upon the acceptance of the paper.

Synthetic Data: Sequences were generated by a self-exciting process with event locations sampled from a mixture of Gaussian distributions which K = 3. Boston Crime: Motor vehicle accident event happened in Boston during 2015.06.14 - 2018.09.03¹ 911 Call: The fire 911 calls reported in Montgomery County, PA, spanning from December 2015 to December 2018, with complete location data.² Earthquake: Earthquakes in the West U.S. with a magnitude of a least 2.5 from 1990 to 2020 were recorded by the U.S. Geological Survey.³

5.2 BASELINE

364 We evaluate our proposed model by comparing it to several spatial models, temporal models, and spatio-temporal point process models across various datasets. Neural Embedding Spatio-Temporal (NEST) Zhu et al. (2021b): Mixture Gaussian diffusion kernels based spatio-temporal process model, 366 learning parameters by imitation learning. Spatio-Temporal Diffusion Point Processes(DSTPP) 367 Yuan et al. (2023): Diffusion model-based framework for spatio-temporal point processes, enabling 368 the learning of complex joint distributions. Neural Spatio-temporal Point Process (NSTPP) Chen 369 et al. (2020): Neural ODE-based STPP model, utilizing Neural Jump SDEs for parameterizing 370 temporal intensity and continuous-time normalizing flows for spatial PDFs. Neural Point Process 371 for Learning Spatiotemporal Event Dynamics (DeepSTPP) Zhou et al. (2022): Variational 372 inference based deep STPP model. Continuous normalizing flow (CNF) Chen et al. (2018): Neural 373 ordinary differential equations based spatial prediction model. Transformer Hawkes Process (THP) 374 Zuo et al. (2020): Transformer based temporal point process model.

375

¹https://www.kaggle.com/datasets/AnalyzeBoston/crimes-in-boston

²https://www.kaggle.com/code/mohammadhy/analysis-911-call

³https://www.usgs.gov/programs/earthquake-hazards



Figure 2: The comparison of exploration efficiency with/without causal action selection.

5.3 EVALUATION METRICS

We assess the models' performance from two angles: (1) Computing mean square error (MSE) for the one-step ahead prediction of temporal and spatial (2) To demonstrate its effectiveness in addressing compounding prediction errors, we calculate the accumulative MSE for 20 steps ahead prediction. We refer readers to Appendix D for more details on the evaluation metrics, parameters setting, and hyperparameter selection experiments.

397 5.4 PERFORMANCE

378

379

380 381

382

384

385

386 387

388 389

390

391

392

393

394

395 396

Table 1 and 2 present the comprehensive performance of models in terms of prediction accuracy and ability to address the compounding prediction error issues, respectively. (See Appendix C for additional performance results on methods that focus exclusively on spatial or temporal aspects.) Additionally, Figure 2 shows the comparison of reinforcement learning exploration efficiency with/without causal learning. Based on these results, we can draw the following conclusions:

403 Our model significantly alleviates compounding prediction errors. The 20-steps ahead prediction 404 MSE results demonstrate our model has superior performance in mitigating the compounding 405 prediction errors. This success is attributed to the reinforcement learning strategy, which can learn 406 robust policies that account for dependencies between sequential decisions. Furthermore, instead 407 of making point predictions at each time step independently, RL learns a policy that maps states 408 to actions, which reduces the errors arising from historical predictions. This approach stands in 409 contrast to MLE and other methods, which do not inherently consider the sequential decision-making 410 process and the potential accumulation of prediction errors. Additionally, our model achieves good 411 performance in one-step prediction, demonstrating that the introduction of reinforcement learning and the focus on multi-step forecasting do not compromise the accuracy of short-term predictions. 412

413 Following actions with the most Causal influence can improve the exploration efficiency. From 414 the convergence experiments depicted in Figure 2, it is obvious that the model exhibits higher 415 exploration efficiency when employing a causal influence-based action selection method. This 416 enhanced efficiency can be attributed to the causal influence-based approach enables the model to prioritize actions that are more likely to influence the outcome of interest directly. By understanding 417 and leveraging the causal relationships within the environment, the model can avoid wasting effort 418 on exploring actions with minimal or no impact on the desired outcome, leading to more efficient 419 exploration. 420

Appropriate parametric assumptions are essential for the effectiveness of point processes. These
 assumptions about the parameter space can significantly influence the model's ability to accurately
 capture the underlying temporal or spatial structures inherent in the data. When the assumptions align
 well with the real-world phenomena being modeled, point processes can offer precise predictions,
 uncover hidden patterns, and provide deep insights into the dynamics of complex systems. Conversely,
 if the parametric assumptions are not well-suited to the data, the model may fail to converge, exhibit
 poor predictive performance, or even misinterpret the causal relationships within the data.

Intensity recovery on synthetic data. To better understand the mixture density network mechanism,
 we examine the location spatial distribution at different stages of the recovery iterations. The figures
 3 illustrate the intensity recovery performance of our models in discrete-space scenarios, Initially,
 the spatial distribution contains 3 distributions with blurred appearance. This stage represents the
 early phase of training, where the separation between regions is still not clear. As the reinforcement



Figure 4: Ablation study on the action selection and reward function.

459 learning process continues, the spatial distribution becomes more concentrated and the noise level is 460 significantly reduced. This suggests that the model is increasingly learning the underlying patterns. In the final stages, the spatial distribution is close to the ground truth, indicating that the model has effectively mastered the generative process of the spatial distribution.

462 463 464

461

457 458

5.5 ABLATION STUDY

465 In our ablation study, we evaluated the impact of two critical components: the causal inference-466 based action selection, and MMD-based reward. The baseline model CRLSTPP, which includes all 467 components, achieved the highest performance metrics. When the MMD reward was modified to the 468 L1 distance-based reward, we observed a significant drop in accuracy, indicating the essential role of 469 the reward design in this RL model. Excluding causal inference-based action selection resulted in a 470 moderate decline in accuracy, this is because the causal inference-based action selection method can 471 improve the exploration efficiency significantly but not influence the accuracy too much.

472 473

6 CONCLUSION

474 475

We propose a novel causal reinforcement learning framework for modeling spatio-temporal event 476 sequences. Under this framework, the policy is designed as a special recurrent mixture density 477 network (RMDN), enabling the sequential sampling of discrete events. We update the policy by 478 maximizing the reward which correlates with the disparity between the generated sequences and 479 observed sequences. This approach effectively reduces the error accumulation in likelihood-based 480 methodologies for multi-step forecasting. Furthermore, to improve the exploration efficiency of our 481 reinforcement learning model, we have implemented a strategy that utilizes causal information to 482 optimize event generation. This is achieved by selecting the event (action) that induces the maximum deviation in mutual information from its expected values. Experiments on synthetic and real-world 483 datasets demonstrated our model's methodological robustness and superior performance in modeling 484 complex spatio-temporal event data. This framework can be extended to accommodate more complex 485 causal structures and explore the integration of additional data types in future works.

486 REFERENCES

- Murad Ali. Imitation or innovation: To what extent do exploitative learning and exploratory learning
 foster imitation strategy and innovation strategy for sustained competitive advantage?. *Technological Forecasting and Social Change*, 165:120527, 2021.
- Kushal Arora, Layla El Asri, Hareesh Bahuleyan, and Jackie Chi Kit Cheung. Why exposure bias
 matters: An imitation learning perspective of error accumulation in language generation. *arXiv preprint arXiv:2204.01171*, 2022.
- Adrian Baddeley, Imre Bárány, and Rolf Schneider. Spatial point processes and their applications. *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004*, pp. 1–75, 2007.
- Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron
 Courville, and Yoshua Bengio. An actor-critic algorithm for sequence prediction. *arXiv preprint arXiv:1607.07086*, 2016.
- Loris Bazzani, Hugo Larochelle, and Lorenzo Torresani. Recurrent mixture density network for spatiotemporal visual attention. *arXiv preprint arXiv:1603.08199*, 2016.
- Ioana Bica, Daniel Jarrett, and Mihaela van der Schaar. Invariant causal imitation learning for
 generalizable policies. *Advances in Neural Information Processing Systems*, 34:3952–3964, 2021.
- Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- Ricky TQ Chen, Brandon Amos, and Maximilian Nickel. Neural spatio-temporal point processes.
 arXiv preprint arXiv:2011.04583, 2020.
- Peter J Diggle. Spatio-temporal point processes: methods and applications. *Monographs on Statistics and Applied Probability*, 107:1, 2006.
- 514 Nan Ding and Radu Soricut. Cold-start reinforcement learning with softmax policy gradient. Advances
 515 *in Neural Information Processing Systems*, 30, 2017.
- Wenhao Ding, Haohong Lin, Bo Li, and Ding Zhao. Generalizing goal-conditioned reinforcement
 learning with variational causal reasoning. *Advances in Neural Information Processing Systems*, 35:26532–26548, 2022.
- ⁵¹⁹ Zheng Dong, Matthew Repasky, Xiuyuan Cheng, and Yao Xie. Deep graph kernel point processes.
 arXiv preprint arXiv:2306.11313, 2023a.
- Zheng Dong, Shixiang Zhu, Yao Xie, Jorge Mateu, and Francisco J Rodríguez-Cortés. Non-stationary
 spatio-temporal point process modeling for high-resolution covid-19 data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(2):368–386, 2023b.
- Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58 (1):83–90, 1971.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.
- Fatih Ilhan and Suleyman S Kozat. Modeling of spatio-temporal hawkes processes with randomized kernels. *IEEE Transactions on Signal Processing*, 68:4946–4958, 2020.
- Amirkasra Jalaldoust, Kateřina Hlaváčková-Schindler, and Claudia Plant. Causal discovery in hawkes
 processes by minimum description length. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 6978–6987, 2022.
- Beomjoon Kim and Joelle Pineau. Maximum mean discrepancy imitation learning. In *Robotics: Science and systems*, 2013.
- Jin Li, Ye Luo, and Xiaowei Zhang. Causal reinforcement learning: An instrumental variable approach. *arXiv preprint arXiv:2103.04021*, 2021a.

E 40

559

565

566

585

340	Shuang Li, Shuai Xiao, Shixiang Zhu, Nan Du, Yao Xie, and Le Song. Learning temporal point
541	processes via reinforcement learning. Advances in neural information processing systems, 31.
542	2018.
543	

- Shuang Li, Lu Wang, Xinyun Chen, Yixiang Fang, and Yan Song. Understanding the spread of
 covid-19 epidemic: A spatio-temporal point process view. *arXiv preprint arXiv:2106.13097*, 2021b.
- 547
 548
 548
 549
 550
 550
 250
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
 251
- Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Efficient reinforcement learning with prior
 causal knowledge. In *Conference on Causal Learning and Reasoning*, pp. 526–541. PMLR, 2022.
- Hongyuan Mei and Jason M Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. Advances in neural information processing systems, 30, 2017.
- George O Mohler, Martin B Short, P Jeffrey Brantingham, Frederic Paik Schoenberg, and George E
 Tita. Self-exciting point process modeling of crime. *Journal of the american statistical association*, 106(493):100–108, 2011.
- Jesper Moller and Rasmus Plenge Waagepetersen. Statistical inference and simulation for spatial point processes. CRC press, 2003.
- Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401):9–27, 1988.
 - Yosihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50:379–402, 1998.
- Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*, 2015.
- Maximilian Seitzer, Bernhard Schölkopf, and Georg Martius. Causal influence detection for improv ing efficiency in reinforcement learning. *Advances in Neural Information Processing Systems*, 34: 22905–22918, 2021.
- Bowen Tan, Zhiting Hu, Zichao Yang, Ruslan Salakhutdinov, and Eric P Xing. Connecting the dots between mle and rl for sequence generation. 2018.
- Lingxiao Wang, Zhuoran Yang, and Zhaoran Wang. Provably efficient causal reinforcement learning
 with confounded observational data. *Advances in Neural Information Processing Systems*, 34: 21164–21175, 2021.
- Shuai Xiao, Mehrdad Farajtabar, Xiaojing Ye, Junchi Yan, Le Song, and Hongyuan Zha. Wasserstein
 learning of deep generative point process models. *Advances in neural information processing* systems, 30, 2017.
- Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. Learning granger causality for hawkes
 processes. In *International conference on machine learning*, pp. 1717–1726. PMLR, 2016.
- Guolei Yang, Ying Cai, and Chandan K Reddy. Recurrent spatio-temporal point process for check-in time prediction. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 2203–2211, 2018.
- Yuan Yuan, Jingtao Ding, Chenyang Shao, Depeng Jin, and Yong Li. Spatio-temporal diffusion point processes. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3173–3184, 2023.
- 593 Yan Zeng, Ruichu Cai, Fuchun Sun, Libo Huang, and Zhifeng Hao. A survey on causal reinforcement learning. *arXiv preprint arXiv:2302.05209*, 2023.

594 595 596	Yixuan Zhang, Quyu Kong, and Feng Zhou. Integration-free training for spatio-temporal multimodal covariate deep kernel point processes. <i>Advances in Neural Information Processing Systems</i> , 36: 25031–25049, 2023.
597 598 599	Zihao Zhou and Rose Yu. Automatic integration for spatiotemporal neural point processes. <i>Advances in Neural Information Processing Systems</i> , 36, 2024.
600 601 602	Zihao Zhou, Xingyi Yang, Ryan Rossi, Handong Zhao, and Rose Yu. Neural point process for learning spatiotemporal event dynamics. In <i>Learning for Dynamics and Control Conference</i> , pp. 777–789. PMLR, 2022.
603 604 605 606	Shixiang Zhu, Ruyi Ding, Minghe Zhang, Pascal Van Hentenryck, and Yao Xie. Spatio-temporal point processes with attention for traffic congestion event modeling. <i>IEEE Transactions on Intelligent Transportation Systems</i> , 23(7):7298–7309, 2021a.
607 608 609	Shixiang Zhu, Shuang Li, Zhigang Peng, and Yao Xie. Imitation learning of neural spatio-temporal point processes. <i>IEEE Transactions on Knowledge and Data Engineering</i> , 34(11):5391–5402, 2021b.
610 611 612 613	Shixiang Zhu, Minghe Zhang, Ruyi Ding, and Yao Xie. Deep fourier kernel for self-attentive point processes. In <i>International Conference on Artificial Intelligence and Statistics</i> , pp. 856–864. PMLR, 2021c.
614 615	Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. Transformer hawkes process. In <i>International conference on machine learning</i> , pp. 11692–11702. PMLR, 2020.
617	
618	
619	
620	
621	
622	
623	
624	
625	
626	
627	
628	
629	
630	
631	
632	
633	
634	
635	
637	
638	
639	
640	
641	
642	
643	
644	
645	
646	
647	

A CAUSAL INFLUENCE DETECTION BASED ON CMI

Conditional Mutual Information (CMI) I(X; Y | Z) measures the amount of information that two random variables X and Y share, given the knowledge of a third variable Z. In the context of reinforcement learning, we can think of X as the action a, Y as the current state s, and Z as the previous state s_{i-1}

CMI is given by the formula:

$$I(S_i; A \mid S_{i-1}) = H(S_i \mid S_{i-1}) - H(S_i \mid S_{i-1}, A)$$

where:

• $H(S_i | S_{i-1})$ is the conditional entropy of S_i given S_{i-1} .

• $H(S_i | S_{i-1}, A)$ is the conditional entropy of S_i given S_{i-1} and A.

To understand the relationship between CMI and the KL divergence used in our action selection process, we delve into the derivation. The KL divergence $D_{KL}(P||Q)$ between two distributions Pand Q measures how one probability distribution diverges from a expected probability distribution. In our context, P is $p(s_i | s_{i-1}, a)$ and Q is the average distribution $\frac{1}{Z} \sum_{k=1}^{Z} p(s_i | s_{i-1}, a^z)$. The KL divergence is expressed as:

$$D_{KL}\left(p(s_i \mid s_{i-1}, a) \middle\| \frac{1}{Z} \sum_{k=1}^{Z} p(s_i \mid s_{i-1}, a^z)\right)$$

This KL divergence measures how much information the action a provides about the state s_i , conditioned on s_{i-1} .

678 CMI can be expressed in terms of KL divergence as follows:

$$I(S_i; A \mid S_{i-1}) = \mathbb{E}_{S_{i-1}} \left[D_{KL}(p(S_i \mid S_{i-1}, A) \| p(S_i \mid S_{i-1})) \right]$$

Here, $p(S_i | S_{i-1})$ is the marginal distribution of S_i given S_{i-1} , which can be approximated by averaging over all actions a.

The action selection criterion aims to maximize the KL divergence, effectively choosing the action that maximizes the deviation from the expected next state:

$$a^* = \arg \max_{a \in \{a^1, \dots, a^z\}} \left(D_{KL} \left(p(s_i \mid s_{i-1}, a) \middle\| \frac{1}{Z} \sum_{k=1}^Z p(s_i \mid s_{i-1}, a^z) \right) \right)$$

This criterion is analogous to selecting the action a that maximizes the CMI $I(S_i; A | S_{i-1})$. By doing so, it chooses the action that provides the most information about the next state given the current state. This approach effectively enhances the exploration efficiency within the reinforcement learning framework by focusing on actions that exert significant causal influence on the environment.

B DATASET

Synthetic: Sequences were generated by a self-exciting process with event locations sampled from a mixture of Gaussian distributions which K = 3. We set base rate =0.5, event influence=1.2, decay rate=0.5, max time=20, number of sequences =100 and S = [-1, 1], for the Gaussian mixture model,

our parameters are as follows:

Means:
$$\mu_1 = \begin{pmatrix} 0 \\ 0.1 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \quad \mu_3 = \begin{pmatrix} -0.5 \\ -0.6 \end{pmatrix}$$

Covariances: $\Sigma_1 = \begin{pmatrix} 0.05 & 0.025 \\ 0.025 & 0.05 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0.05 & -0.025 \\ -0.025 & 0.055 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 0.075 & 0.025 \\ 0.025 & 0.05 \end{pmatrix}$
Weights: $w_1 = 0.35, \quad w_2 = 0.35, \quad w_3 = 0.3$

(-0.5)

It contains 100 sequences and the average length of sequences is 128.

Boston Crime: The Boston dataset, sourced from Kaggle, encompasses sequences of 65 distinct types of crime incidents, including burglary, street crime, and auto theft, gathered over 3 years from the Boston Police Department. Notably, the category "Motor vehicle accident" is what we use in this experiment. Each recorded crime event is marked by a timestamp, location, and other information. After deleting some abnormal values and nan values, we cut it into 168 sequences by time, the average length of sequences is 221.

911 Call: The 911Call dataset comprises sequences of emergency phone call records, encompassing EMS, fire, and traffic incidents reported in Montgomery County, PA, spanning from December 2015 to December 2018, with complete location coordinate data. We utilize the category "fire" and cut it into 1100 sequences by time, the average length of these sequences is 91.

Earthquakes: Earthquakes in the West US (Latitude from 32.000 to 49.000, Longitude from -125.000 to -112.000) from 2000 to 2020 with a magnitude of at least 3.0 are collected from the U.S.Geological Survey. Each record has the magnitude, timestamp, and location coordinate. After deleting some abnormal values, we cut it into 120 sequences by month, the average length of sequences is 89.

С ADDITIONAL RESULT

Table 3: Evaluation of performance in predicting time of the next 1 event.									
_		Synthetic		Earthquake		911 Call		Crime	
	Model	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial
	Hawkes	0.327	-	1.010	-	0.871	-	1.432	-
	NHP	0.165	-	0.415	-	0.227	-	0.603	-
	THP	0.161	-	0.412	-	0.230	-	0.601	-
(CRLSTPP	0.152	0.089	0.393	6.608	0.197	0.182	0.581	0.127

Table 4: Evaluation of performance in predicting time of the next 20 event.

	Synthetic		Earthquake		911 Call		Crime	
Model	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial	Temporal	Spatial
Hawkes	8.216	-	19.029	-	28.235	-	39.245	-
NHP	4.621	-	12.091	-	8.926	-	16.144	-
THP	4.347	-	11.127	-	8.014	-	15.924	-
CRLSTPP	3.293	2.017	8.024	201.283	7.021	3.713	13.038	3.282

Hawkes ProcessHawkes (1971): a self-exciting point process where each event increases the likelihood of future events occurring in the near term.

The Neural Hawkes Process (NHP)Mei & Eisner (2017): a LSTM based Hawkes process model.

IMPLEMENTATION DETAILS D

We adopt different metrics to evaluate the models. Firstly, we split the sequence to 80% training and 20% test and we evaluate the test set performance. Our task is focusing on the prediction on

756	short term and long term (which will always cause the compounding prediction error issue) In this
757	setting, we will use events before time t to predict 1 step or 20 steps after t in the test set. We use
758	MSE to evaluate the results of temporal prediction and spatial prediction, respectively. The details of
759	hyper-parameters used in training are the following. We set K=3.5.5.3 for synthetic, Boston, 911Call,
760	and Earthquake, respectively. The detail hyperparameter selection experiment shown in Table 5. In
761	addition, for most of the datasets, we set sampling number=10, learning rate=2e-4, head number=4,
762	layer number=4, and model dimension $D = 512$. Our experiment resource including 2 NVIDIA RTX
763	A5000 GPU cards.

Table 5: Hyperparameter selection of Gaussian component number K on synthetic data.

		-	
Model	K=1	K=3	K=5
Temporal - 1 Steps	0.160	0.152	0.154
Spatial - 1 Steps	1.002	0.089	0.093
Temporal - 20 Steps	3.472	3.293	3.409
Spatial - 20 Steps	3.007	2.023	2.825