

Tsz Nam Chan  
Shenzhen University  
edisonchan@szu.edu.cn

Bojian Zhu  
Hong Kong Baptist University  
csbjzhu@comp.hkbu.edu.hk

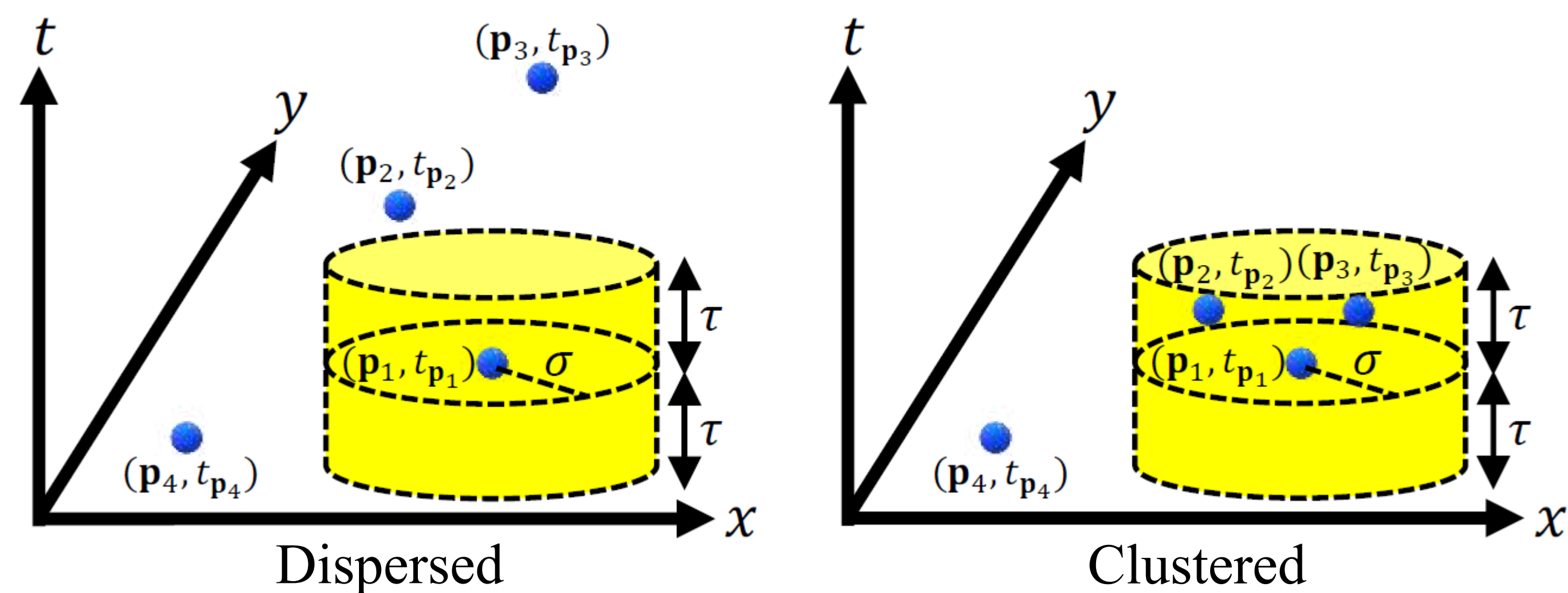
Dingming Wu  
Shenzhen University  
dingming@szu.edu.cn

Renchi Yang  
Hong Kong Baptist University  
renchi@hkbu.edu.hk

Ruisheng Wang  
Shenzhen University  
ruiswang@szu.edu.cn



## What is Spatiotemporal K-function-based Analysis?



- Spatiotemporal K-function counts those data points in  $P$  (with size  $n$ ) that are within the spatial threshold  $\sigma$  and the temporal threshold  $\tau$  from each data point  $(\mathbf{p}_i, t_{\mathbf{p}_i})$ .

$$K_P(\sigma, \tau) = \sum_{(\mathbf{p}_i, t_{\mathbf{p}_i}) \in P} \sum_{\substack{(\mathbf{p}_j, t_{\mathbf{p}_j}) \in P \\ j \neq i}} \mathcal{I}(d(\mathbf{p}_i, \mathbf{p}_j) \leq \sigma, d(t_{\mathbf{p}_i}, t_{\mathbf{p}_j}) \leq \tau)$$

- Domain experts need to (1) determine  $S$  spatial thresholds,  $\sigma_1, \sigma_2, \dots, \sigma_S$ , and  $T$  temporal thresholds,  $\tau_1, \tau_2, \dots, \tau_T$ , (2) generate  $L$  random datasets (with the same size  $n$ ),  $R_1, R_2, \dots, R_L$ , and (3) compute the following spatiotemporal K-function values for each  $(\sigma, \tau)$ -pair:

- $K_P(\sigma, \tau)$  (the green surface)
- $\mathcal{L}(\sigma, \tau) = \min(K_{R_1}(\sigma, \tau), K_{R_2}(\sigma, \tau), \dots, K_{R_L}(\sigma, \tau))$  (the red surface)
- $\mathcal{U}(\sigma, \tau) = \max(K_{R_1}(\sigma, \tau), K_{R_2}(\sigma, \tau), \dots, K_{R_L}(\sigma, \tau))$  (the blue surface)

- Applications:

- Obtain the suitable parameters for some models, e.g., clustering/hotspot detection.
- Determine whether the analysis results (e.g., COVID-19 hotspots) are meaningful.



(a) 24th December 2022

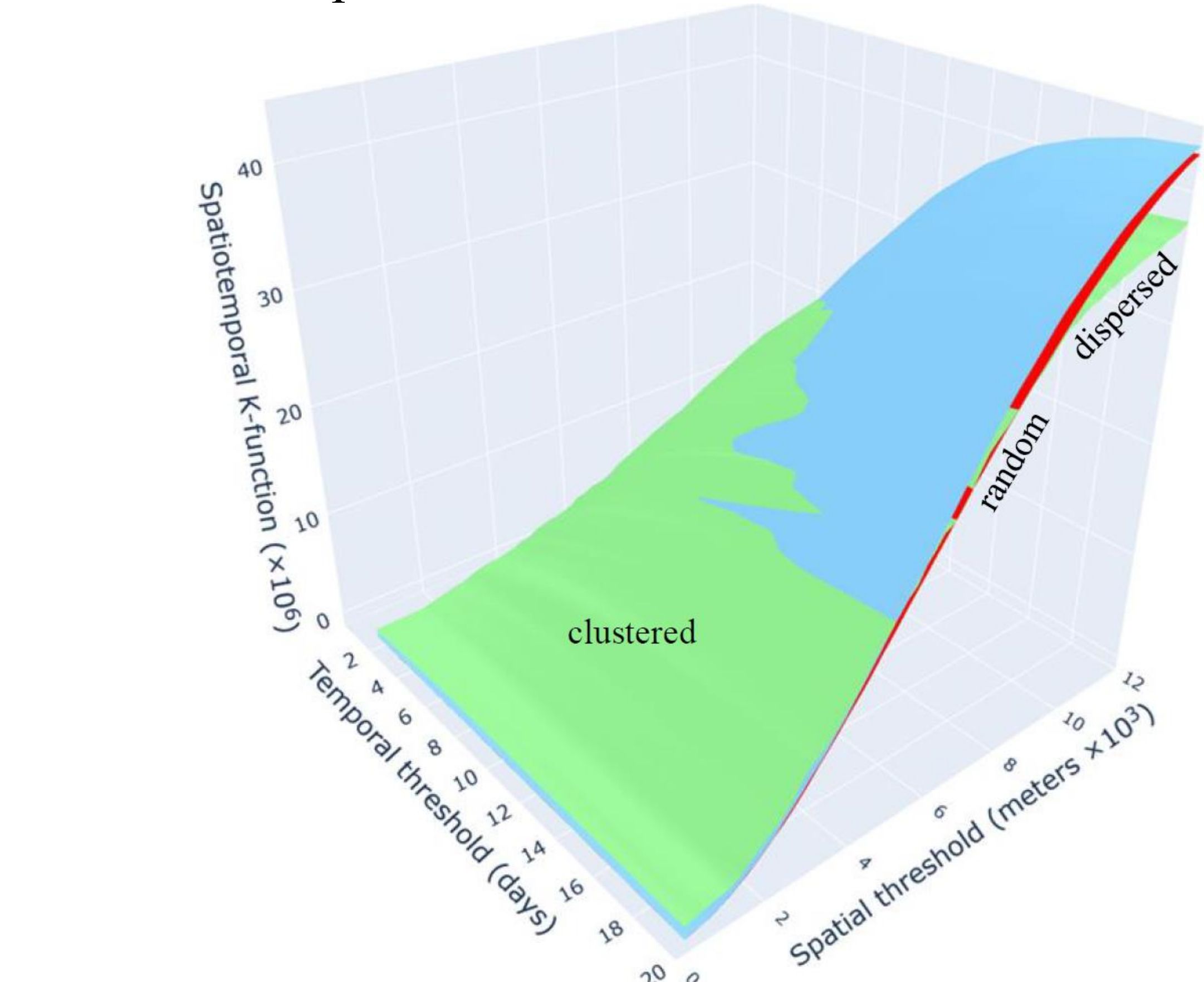
(b) 29th December 2022

(c) 3rd January 2023

(d) 8th January 2023

Spatiotemporal Kernel Density Visualization (Hong Kong COVID-19 cases)

Spatiotemporal K-function plot (Hong Kong COVID-19 cases)



## A Distribution and Aggregation Solution (DNA)

- Representation of K-function by the range query set  $\mathbb{R}_{\sigma, \tau}(\mathbf{p}_i, t_{\mathbf{p}_i})$  for each data point  $(\mathbf{p}_i, t_{\mathbf{p}_i})$ .

$$K_P(\sigma, \tau) = \sum_{(\mathbf{p}_i, t_{\mathbf{p}_i}) \in P} |\mathbb{R}_{\sigma, \tau}(\mathbf{p}_i, t_{\mathbf{p}_i})|$$

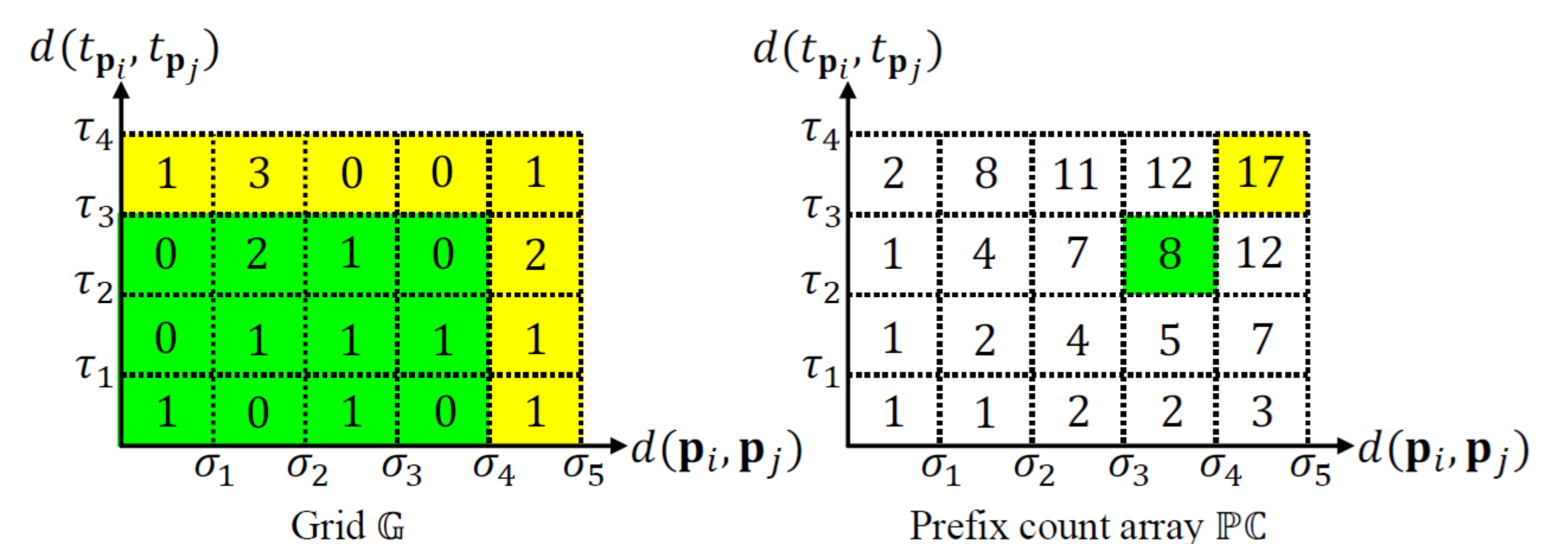
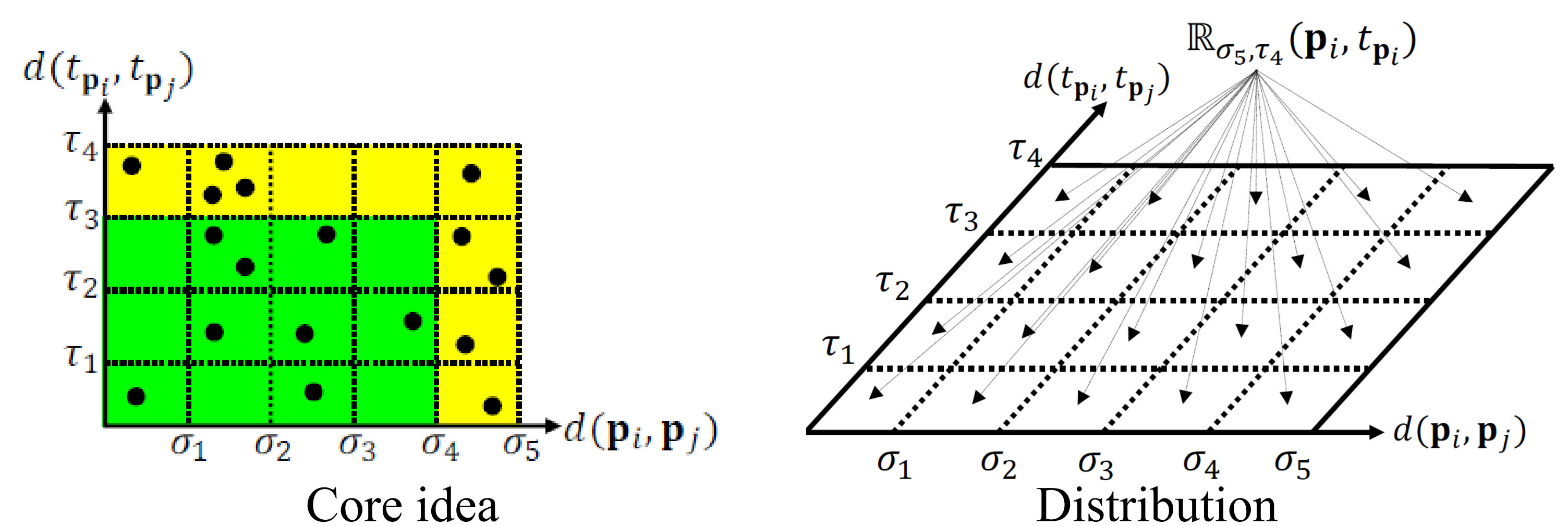
where

$$\mathbb{R}_{\sigma, \tau}(\mathbf{p}_i, t_{\mathbf{p}_i}) = \left\{ (\mathbf{p}_j, t_{\mathbf{p}_j}) \in P : d(\mathbf{p}_i, \mathbf{p}_j) \leq \sigma, d(t_{\mathbf{p}_i}, t_{\mathbf{p}_j}) \leq \tau \right\}$$

- The range query set  $\mathbb{R}_{\sigma_S, \tau_T}(\mathbf{p}_i, t_{\mathbf{p}_i})$  with the largest spatial threshold  $\sigma_S$  and the largest temporal threshold  $\tau_T$  covers the range query set  $\mathbb{R}_{\sigma_u, \tau_v}(\mathbf{p}_i, t_{\mathbf{p}_i})$  with each  $(\sigma_u, \tau_v)$ -threshold pair.

$$\mathbb{R}_{\sigma_u, \tau_v}(\mathbf{p}_i, t_{\mathbf{p}_i}) \subseteq \mathbb{R}_{\sigma_S, \tau_T}(\mathbf{p}_i, t_{\mathbf{p}_i})$$

- Time Complexity:  $O(Ln^2 + LSTn)$  ☺
- Distribution:  $O(n + ST)$  for each data point.
- Aggregation:  $O(ST)$  for each data point.

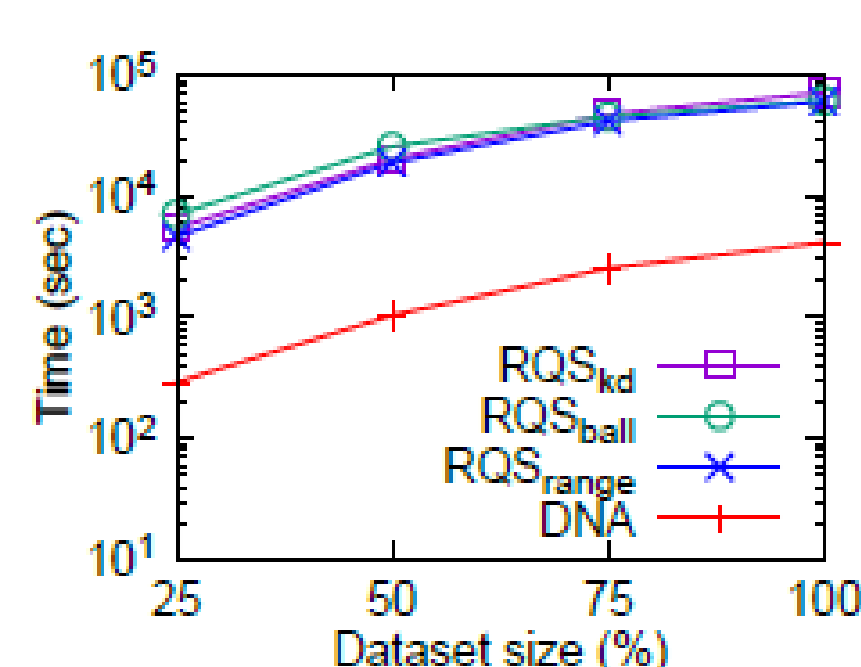


Aggregation

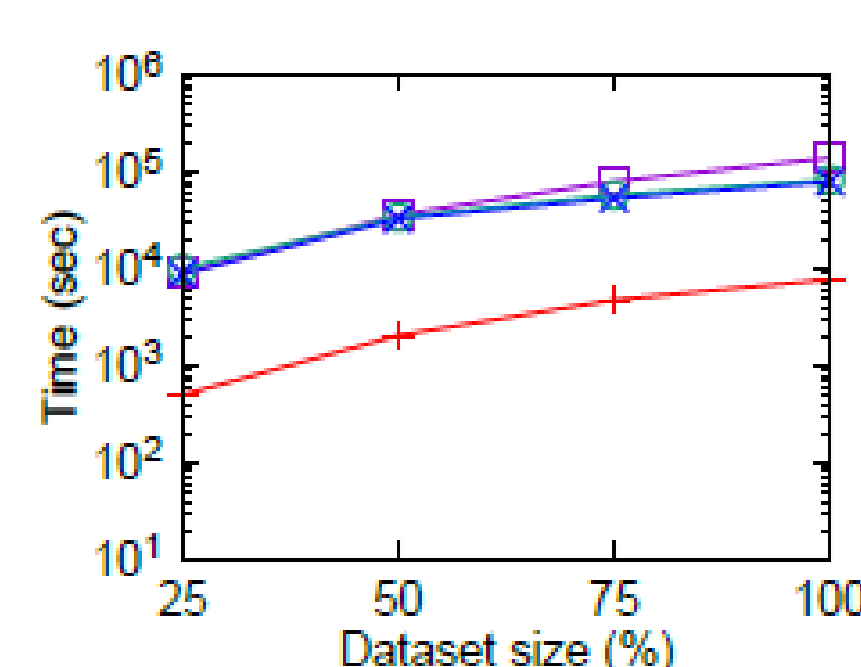
## Theoretical and Experimental Results

Threshold interval type	Method	Time complexity	Space complexity
Fixed threshold intervals	RQS	$O(LSTn^2)$	$O(nL + ST)$
	DNA	$O(Ln^2 + LSTn)$ (Theorem 1)	$O(nL + ST)$ (Theorem 3)
Multiple threshold intervals	RQS	$O(LSTn^2)$	$O(nL + ST)$
	DNA	$O(Ln^2 \log ST + LSTn)$ (Theorem 2)	$O(nL + ST)$ (Theorem 3)

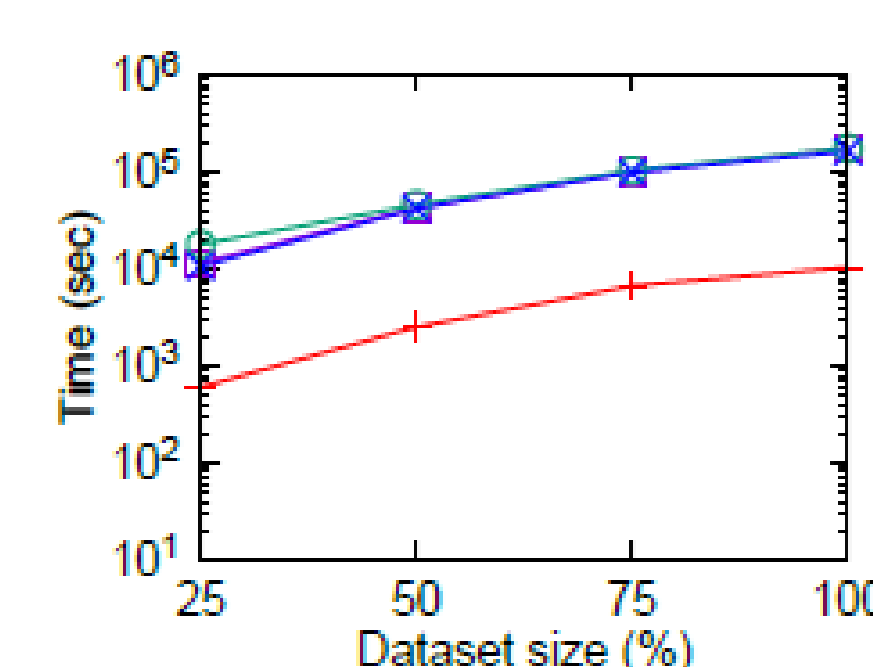
Name	$n$	Category
Seattle	236,136	Traffic accidents
Atlanta	363,861	Crime events
Los Angeles	559,861	Bicycle mobility
New York	1,541,284	Traffic accidents



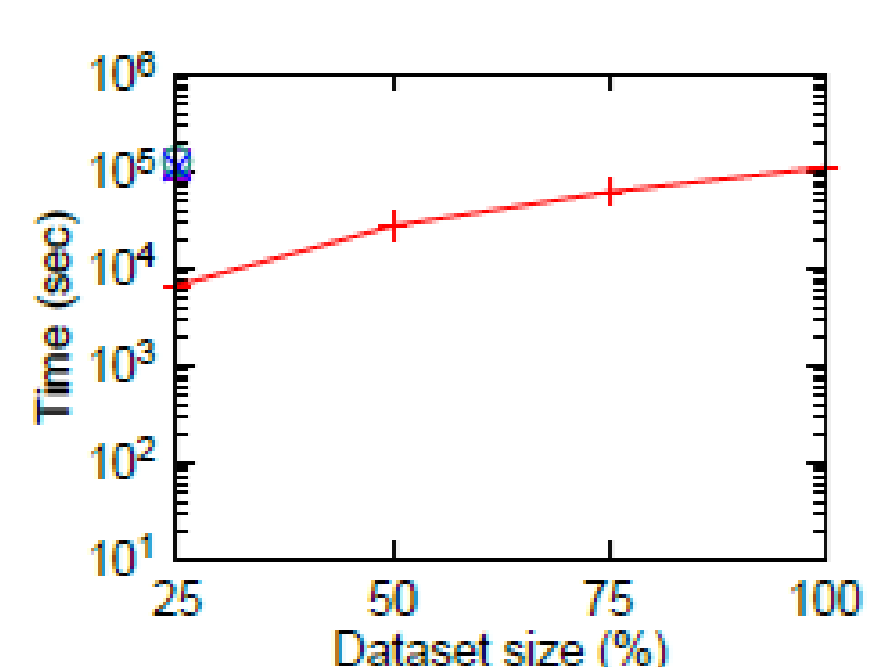
(a) Seattle



(b) Atlanta



(c) Los Angeles



(d) New York