Space-Time Kernel Density Estimation

I. Introduction

The rapid propagation of infectious diseases (e.g. zika, Ebola, H1N1, dengue fever) is conducive to serious, epidemic outbreaks, posing a threat to vulnerable populations. Such diseases have complex transmission cycles, and effective public health responses require the ability to monitor outbreaks in a timely manner [EE11]. Space-time statistics facilitate the discovery of disease dynamics including rate of spread, seasonal cyclic patterns, direction, intensity (i.e. clusters), and risk of diffusion to new regions. However, obtaining accurate results from space-time statistics is computationally very demanding, which is problematic when public health interventions are promptly needed. The issues of computational efforts are exacerbated with spatiotemporal datasets of increasing size, diversity and availability [GWM14]. High-performance computing reduces the effort required to identify these patterns, however heterogeneity in the data must be accounted for [HDTC16].

Epidemiology is only one of the application domains where it is important to understand how events occurring at different locations and different times form clusters. Political analysis, social media analysis, or the study of animal migration also require the understanding of spatial and temporal locality of events. The massive amount of data we see nowadays is often analyzed using complex models. But before these can be constructed, data scientists need to interactively visualize and explore the data to understand its structure.

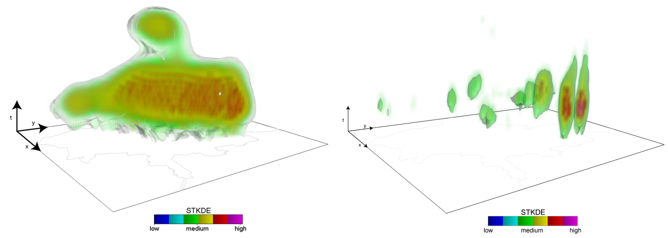
II. Space-Time Kernel Density Estimation

A. Description

Space-time kernel density (STKDE) is used for identifying spatiotemporal patterns in datasets. It is a temporal extension of the traditional 2D kernel density estimation [Sil86] which generates density surface ("heatmap") from a set of n points located in a geographic space. The resulting density estimates are visualized within the space-time cube framework using two spatial (x, y) and a temporal dimension (t) [NY10]. STKDE creates a discretized 3D volume where each voxel (3D equivalent of a pixel) is assigned a density estimate based on the surrounding points. The space-time density is estimated using (following the notations of [HDTC16]):

$$\hat{f}(x, y, t) = \frac{1}{nh_s^2 h_t} \sum_{i|d_i<h_s, t_i<h_t} k_s\left(\frac{x - x_i}{h_s}, \frac{y - y_i}{h_s}\right) k_t\left(\frac{t - t_i}{h_t}\right)$$

Density $\hat{f}(x, y, t)$ of each voxel is determined by number and distance of events (points) (x_i, y_i, t_i) within its vicinity, which is conceptualized by a cylinder. The spatial bandwidth



(a) $h_s = 2500m$, $h_t = 14days$ (b) $h_s = 500m$, $h_t = 7days$

Fig. 1. Visualization of Dengue fever cases in Cali, Colombia in 2010 and 2011 for different spatial bandwidth and temporal bandwidth.

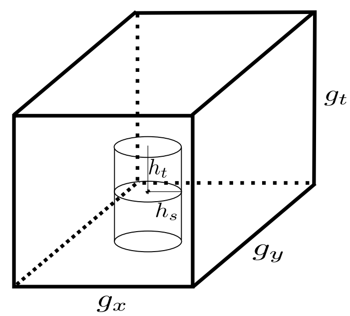


Fig. 2. The computation of STKDE happens in a domain space of size g_x, g_y, g_t. Each point impacts the neighboring space at a distance h_s in space and h_t in time; forming a cylinder of diameter $2h_s$ and of height $2h_t$.

h_s forms a circle which, due to the orthogonal relationship between space and time, is extended to a cylinder by temporal bandwidth h_t. For every event inside the cylinder, the spatial (d_i) and temporal (t_i) distances are smaller than h_s and h_t. Therefore, the event receives a weight based on the kernel functions k_s and k_t (distance decay):

$$k_s(u, v) = \frac{\pi}{2}(1 - u)^2(1 - v)^2$$

$$k_t(w) = \frac{3}{4}(1 - w)^2$$

Figure 1 illustrates how varying the bandwidths used in STKDE helps focusing the graphical visualization of Dengue fever cases in Cali, Colombia [DDC+14]. Figure 2 shows the impact of a single point on the neighboring space.

Computationally, the domain of size g_x, g_y, g_t is discretized in voxels using a spatial resolution $sres$ and a temporal resolution $tres$. Therefore, the domain is represented by a grid of size $G_x = \lceil \frac{g_x}{sres} \rceil$, $G_y = \lceil \frac{g_y}{sres} \rceil$, $G_t = \lceil \frac{g_t}{tres} \rceil$. Each point causes a density increase in the voxels that are within a cylinder centered on the point, of radius equal to the spatial

n	Number of points
$s = (x, y, t)$	A voxel and sampling coordinate
(x_i, y_i, t_i)	Coordinate of point i
h_s, h_t	Spatial and temporal bandwidth
g_x, g_y, g_t	Real size of the domain (in meters)
$sres$, $tres$	Resolution (in meters)
$s = (X, Y, T)$	A voxel in voxel space
(X_i, Y_i, T_i)	Voxel of point i
G_x, G_y, G_t	Size of the domain (in voxels)
H_s, H_t	Bandwidth (in voxels)

TABLE I
NOTATIONS

bandwidth in voxels $H_s = \lceil \frac{h_s}{sres} \rceil$, and of half height equal to the temporal bandwidth $H_t = \lceil \frac{h_t}{tres} \rceil$.

All the notations are summarized in Table I. As a convention all uppercase notations denote quantities in voxels and all lowercase notations denote quantities in domain space.

B. Gold Standard Implementation

The gold standard implementation follows the exact definition of STKDE as it is given above. It is a voxel-based algorithm we call VB. For each voxel, VB finds the points within the temporal and spatial bandwidths and calculates the contribution to the density of this voxel. The pseudo code is given in Algorithm 1.

Algorithm 1 VB

```
for all voxels s = (x, y, t) do
    sum = 0
    for all points i at x_i, y_i, t_i do
        if sqrt((x_i − x)^2 + (y_i − y)^2) < h_s and |t_i − t| ≤ h_t then
            sum+ = k_s((x−x_i)/h_s, (y−y_i)/h_s) k_t((t−t_i)/h_t)
    stkde[X][Y][T] = sum / (n h_s^2 h_t)
```

The algorithm performs $\theta(G_x G_y G_t n)$ distance tests and computes $\theta(n H_s^2 H_t)$ densities. Since H_s is smaller than G_x and G_y and since H_t is smaller than G_t, the complexity of the algorithm is $\theta(G_x G_y G_t n)$ and it requires $\theta(G_x G_y G_t)$ memory.

REFERENCES

[DDC+14] E. Delmelle, C. Dony, I. Casas, M. Jia, and W Tang. Visualizing the impact of space-time uncertainties on dengue fever patterns. *International Journal of Geographical Information Science*, 28(5):1107–1127, 2014.

[EE11] Lars Eisen and Rebecca J Eisen. Using geographic information systems and decision support systems for the prediction, prevention, and control of vector-borne diseases. *Annual review of entomology*, 56:41–61, 2011.

[GWM14] Tony H Grubesic, Ran Wei, and Alan T Murray. Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense. *Annals of the Association of American Geographers*, 104(6):1134–1156, 2014.

[HDTC16] A. Hohl, E. Delmelle, W. Tang, and I. Casas. Accelerating the discovery of space-time patterns of infectious diseases using parallel computing. *Spatial and Spatio-temporal Epidemiology*, 2016.

[NY10] T. Nakaya and K. Yano. Visualising crime clusters in a space-time cube: an exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, 14(3):223–239, 2010.

[Sil86] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.