PrivSTD: Differentially Private Spatio-temporal Trajectory Density Data Publication

Shuzhan Ye
Zhejiang University
yeshuzhan123@zju.edu.cn

Yujia Hu
Zhejiang University
yujiahuzju@gmail.com

Lu Chen
Zhejiang University
luchen@zju.edu.cn

Yangyang Wu
Zhejiang University
zjuwuyy@zju.edu.cn

Zhikun Zhang
Zhejiang University
zhikun@zju.edu.cn

Tianyi Li
Aalborg University
tianyi@cs.aau.dk

Christian S. Jensen
Aalborg University
csj@cs.aau.dk

ABSTRACT

Spatio-temporal trajectory density data is widely used in urban analytics, mobility studies, and epidemiology. While differential privacy enables the release and analysis of such data, providing useful privacy guarantees for high-resolution density data remains challenging. Existing approaches typically perturb density data in the spatial domain, making it difficult to preserve spatio-temporal correlations and causing severe utility loss at fine granularity. We propose PrivSTD, a differentially private method for high-resolution spatio-temporal density release. PrivSTD applies the Discrete Cosine Transform to transform density data into the frequency domain, where low-frequency components capture spatial correlations and temporal smoothness. To suppress noise-dominated frequencies, we introduce a Benjamini–Hochberg FDR-based adaptive truncation mechanism that preserves statistically significant structure at no additional privacy cost. PrivSTD further employs a control variate-enhanced Recorrupted-to-Recorrupted denoising model to reconstruct high-quality density estimates. Experiments show that PrivSTD consistently outperforms existing methods, achieving 1.12×–53.89× lower error (6.12× on average) across all datasets.

PVLDB Reference Format:
Shuzhan Ye, Yujia Hu, Lu Chen, Yangyang Wu, Zhikun Zhang, Tianyi Li, and Christian S. Jensen. PrivSTD: Differentially Private Spatio-temporal Trajectory Density Data Publication. PVLDB, 19(10): 2713 - 2726, 2026.
doi:10.14778/3828612.3828626

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/YeShuzhan/PrivSTD.

1 INTRODUCTION

Large-scale spatiotemporal trajectory data, collected by location-sensing devices such as smartphones and GPS sensors, have become indispensable in domains including urban planning [22, 48], epidemiology [10, 29, 30], and human mobility studies [14, 53]. In many such applications, the release target is not individual trajectories, but aggregated spatiotemporal density data, typically represented as counts over spatiotemporal grids. For example, initiatives such as "Data for Good" [20] have released public spatiotemporal density data to support epidemic tracking [13, 24, 42].

Releasing such data, however, may still expose sensitive individual information. Depending on the application, the privacy risk may concern whether a particular spatiotemporal event is present in the dataset, such as a call record or a location report at a given time, or whether an individual's trajectory and broader activity pattern can be inferred from the released data. These different protection targets give rise to different privacy requirements [2, 4]. Event-level privacy typically targets individual spatiotemporal events generated by sparse or occasional interactions, such as call records [2] or health reports [18] at specific times. By contrast, user-level privacy protects the full trajectory of a user [4]. The latter provides stronger protection, but also induces substantially higher sensitivity.

Existing work on private release of spatiotemporal density data mainly considers event-level privacy. Histogram-based methods [21, 35, 36, 50] and attribute distribution modeling approaches [9, 28, 49] are primarily studied under event-level privacy, while more recent methods such as VDR [4] extend the setting to user-level privacy by combining bounded user contributions with learning-based reconstruction. Despite protecting different targets, both settings face the same utility challenges in high-resolution spatiotemporal density release, and these challenges become more severe under user-level privacy due to its higher sensitivity. Figure 1 uses the real-world CABS_SF dataset to illustrate them.

Challenge I: How to reconcile grid resolution and data utility under differential privacy? Density analysis of spatiotemporal data typically relies on time slicing and grid partitioning. Grid resolution is critical for capturing both local patterns and global mobility trends. As illustrated in Figure 1, low resolution (3×3) yields only coarse distributions, whereas high resolution (32×32) better approximates the original data. However, increasing the resolution drastically dilutes trajectory counts per cell, creating a sparse regime in which DP noise overwhelms the signal. Consequently, direct grid-based

Proceedings of the VLDB Endowment, Vol. 19, No. 10 ISSN 2150-8097.
doi:10.14778/3828612.3828626

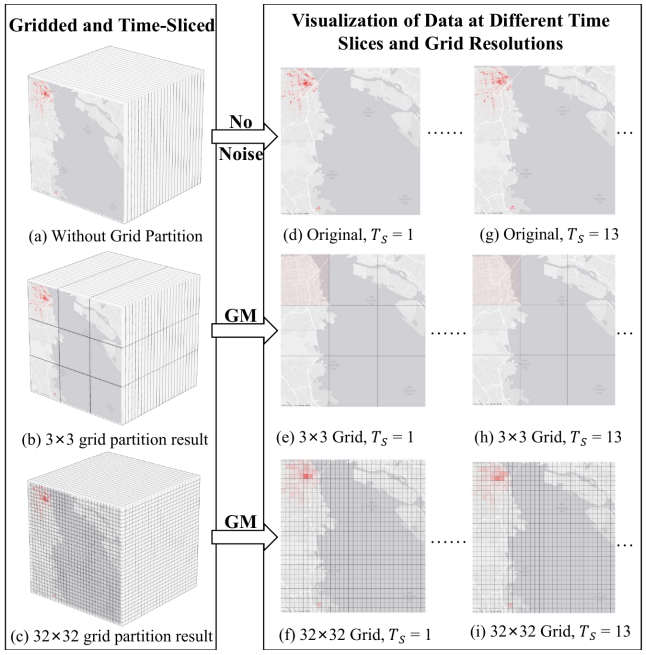

Figure 1: Motivating Example.

perturbations suffer severe utility loss as resolution increases. Motivated by prior transform-based methods [38, 44], PrivSTD uses the Discrete Cosine Transform (DCT) to map density data into the frequency domain. Unlike the spatial domain, where information is dispersed, the DCT concentrates the signal into a compact set of low-frequency coefficients, thereby preserving core features under high-resolution settings.

Challenge II: How to capture spatial and temporal correlations in trajectory density data? As shown in Figure 1(a) and (b), hotspots in both the first time slice ($T_S = 1$) and the 13th time slice ($T_S = 13$) concentrate around similar regions. This reflects strong spatial autocorrelation in trajectory density data, where adjacent grid cells exhibit similar densities due to road networks and continuous movement, as well as temporal smoothness arising from predictable daily and weekly mobility rhythms. However, existing methods do not fully exploit these characteristics in high-resolution spatiotemporal density release. Nearly all current differential privacy mechanisms perturb trajectory counts independently across grid cells, resulting in fragmented spatiotemporal patterns and diminished utility. To overcome this limitation, PrivSTD aggregates the spectral energy of corresponding frequencies across all time slices into a cumulative statistic, denoted by $S(u, v)$. Since persistent spatiotemporal patterns (e.g., recurring hotspots and smooth spatial and temporal variations) appear as stable high-energy components in $S(u, v)$ that are distinct from random noise, we apply the Benjamini–Hochberg False Discovery Rate (BH-FDR) procedure to identify and retain significant coefficients.

Challenge III: How to achieve effective denoising without clean ground-truth data under DP constraints? Under differential privacy constraints, supervised denoising is infeasible because clean ground-truth data are unavailable, forcing it to rely on self-supervised or unsupervised objectives. Existing approaches (e.g., VQ-VAE in VDR, plain R2R, or score matching), however, often degenerate into near-identity mappings that overfit structured and heteroscedastic noise, especially in high-resolution spatiotemporal settings. High dimensionality and varying noise scales across frequency bands further exacerbate training instability. To address this, we introduce the Control Variate-Enhanced Recorrupted-to-Recorrupted (CV-R2R) denoising model, which turns unsupervised denoising into a variance-reduced self-supervised framework that admits a bias-controlled approximation to supervised denoising without access to clean data. By incorporating optimally weighted control variates together with weighted generalized least squares (GLS) and a control-variate-aware Stochastic Polyak Step-size (CV-SPS), CV-R2R improves training stability and reconstruction quality under severe heteroscedastic privacy noise.

To address these challenges, we propose PrivSTD, a differentially private algorithm for high-resolution spatiotemporal density release. Our contributions are summarized as follows.

- We propose PrivSTD, a method for releasing spatiotemporal trajectory density data under differential privacy while preserving spatiotemporal features (Section 4).
- We combine a DCT-based mechanism with BH-FDR-controlled adaptive frequency-domain truncation to alleviate noise amplification in high-resolution scenarios and capture spatiotemporal characteristics across time slices (Section 5).
- We present CV-R2R, a covariance-aware denoising model for structured and heteroscedastic privacy noise. It combines control variates, weighted GLS, and variance-aware Polyak updates, yielding variance reduction for the adopted self-supervised objective and a bias-controlled approximation to supervised denoising (Section 6).
- We report extensive experiments on real-world spatiotemporal trajectory density datasets, demonstrating PrivSTD's effectiveness and efficiency across multiple metrics (Section 7).

2 RELATED WORK

Private Data Release. Since a spatiotemporal density grid is a special case of a multidimensional histogram, our work is closely related to differentially private histogram publishing. Existing methods can be broadly categorized into optimization-based, synthesis-based, and learning-based approaches. Optimization-based methods, such as AHP [50] and MWEM [21], reduce error through adaptive grouping or iterative refinement. Synthesis-based methods, such as PrivBayes [49], use generative models to approximate the underlying data distribution. Learning-based methods capture richer spatiotemporal dependencies, including approaches that decompose data into independent temporal series [2], apply deep generative denoising [4], or approximate query answers via neural databases [47]. However, these methods often restrict query expressiveness, require predefined workloads, or remain sensitive to high-dimensional noise. In particular, methods such as VDR [4] still suffer noticeable utility degradation in high-resolution spatiotemporal density release due to amplified perturbation and limited use of cross-temporal structure.

Transformation-based Utility Enhancement. Compared with direct perturbation in the original domain, perturbing suitably transformed coefficients can improve utility in differentially private data release. PASTE [38] does so for distributed time-series aggregation by compressing recurring query-answer sequences in the Fourier

domain before perturbation and reconstructing them afterward, together with a distributed noise-generation mechanism for the no-trusted-server setting. Privelet [44] applies a transform-then-perturb pipeline to histogram publication over relational data by transforming the frequency matrix into a wavelet domain, perturbing the transformed coefficients, and reconstructing a noisy matrix. Rather than using the transformed representation only as a perturbation space, PrivSTD further exploits spatiotemporal structure through cross-time spectral aggregation and adaptive truncation.

Denoising and Learning under Differential Privacy. Most deep learning-based denoising methods rely on clean–noisy pairs or multiple noisy observations [23, 26, 27, 32], assumptions generally incompatible with differential privacy because repeated access to the same raw data is typically not allowed. While some unsupervised techniques can mitigate low-level noise [32, 52], denoising under DP constraints remains relatively under-explored. Recent work such as VDR [4] has begun to address this issue through regularized representation learning tailored to spatiotemporal density data while maintaining formal privacy guarantees. More broadly, differential privacy has also been incorporated into empirical risk minimization [8, 25] and deep neural networks [1, 41], typically through output perturbation [43], regularization [8, 25], or gradient perturbation [1]. In contrast, methods such as VDR [4] sanitize the data representation before training, thereby decoupling privacy protection from the learning algorithm. Nevertheless, existing unsupervised approaches often degenerate into near-identity mappings that overfit structured and heteroscedastic noise, particularly in high-resolution settings [11, 26].

3 PRELIMINARIES

3.1 Differential Privacy

Differential privacy (DP) [16] is a framework designed for settings involving a trusted data curator. Intuitively, it ensures that the participation of any protected unit has only a limited effect on the output distribution. We adopt *approximate DP* [16], which augments the privacy parameter ϵ with a failure probability δ.

Definition 1. *(ε, δ)-Differential Privacy. An algorithm $\mathcal{A}$ satisfies (ε, δ)-DP if, for any two neighboring datasets D and D', and any measurable subset $T \subseteq \mathrm{Range}(\mathcal{A})$,*

$$\Pr[\mathcal{A}(D) \in T] \leq e^{\epsilon} \Pr[\mathcal{A}(D') \in T] + \delta.$$

Gaussian Mechanism (GM) [51]. Given a query or aggregation function f, the Gaussian mechanism defines a randomized algorithm $\mathcal{A}_f$ that achieves (ϵ, δ)-DP by adding Gaussian noise to $f(D)$. The required noise level is determined by the global sensitivity

$$GS_f = \max_{D \simeq D'} \|f(D) - f(D')\|_2,$$

where D and D' are neighboring datasets under the adopted adjacency relation. For scalar-valued f, the Gaussian mechanism takes the form

$$\mathcal{A}_f(D) = f(D) + \mathcal{N}(0, \sigma^2),$$

where $\mathcal{N}(0, \sigma^2)$ denotes a Gaussian random variable with standard deviation $\sigma \geq \frac{\sqrt{2\ln(1.25/\delta)}\, GS_f}{\epsilon}$. For vector-valued f, $\mathcal{A}_f$ adds independent samples from $\mathcal{N}(0, \sigma^2)$ to each coordinate.

Table 1: Summary of Notation

ϵ	Privacy budget
δ	Failure probability
H, Q	Spatial grid dimensions
T	Number of time slices
U, V	Retained frequency bandwidths
X	Grid-based count/density matrix
Θ	2D DCT coefficient matrix
C_x, C_y	Orthonormal DCT basis matrices
$S(u, v)$	Chi-square statistic used in BH-FDR
W	Weight matrix
G	Covariance factor induced by retained DCT bases
α	Recorruption scaling parameter
y	Private reconstructed density estimate
y^+, y^-	Recorrupted input-target pair
$\mathcal{L}(\theta)$	Basic self-supervised loss
$\mathcal{L}_W(\theta)$	Weighted self-supervised loss
C	Quadratic control variate
$R_W(\theta)$	Weighted supervised risk
K_W	Weighted noise-floor term
δ_W	Alignment bias term

3.2 Privacy Notions and Adjacency

Event-level. Under event-level privacy, an event refers to a user's location report at a timestamp, contributing a single count to one cell of the aggregated spatiotemporal histogram. Equivalently, each record is treated as if it were contributed by a distinct user. Two datasets D and D' are event-level neighbors, denoted $D \simeq_{\text{event}} D'$, if they differ by the addition or removal of a single report.

User-level. Under user-level privacy, the protected unit is an entire user together with all records contributed by that user. Two datasets D and D' are user-level neighbors, denoted $D \simeq_{\text{user}} D'$, if they differ by the addition or removal of all records of a single user. Compared with event-level privacy, user-level privacy provides stronger protection but incurs substantially higher sensitivity.

3.3 Problem Formulation

We consider a dataset D of user location records $(\text{lon}, \text{lat}, \text{time}, \text{id})$ and aim to release a high-resolution spatiotemporal histogram under differential privacy. Space and time are discretized into a three-dimensional histogram $H \in \mathbb{R}^{M \times M \times T}$, to which each report contributes one count in its corresponding spatiotemporal cell. Under either privacy notion, the release target remains the same aggregated spatiotemporal histogram, while the protected unit may be either a single spatiotemporal report or an entire user. In the main paper, we present PrivSTD under the event-level setting; extending it to user-level privacy mainly requires per-user contribution bounding and corresponding sensitivity recalibration.

We design a mechanism $\mathcal{M}$ that takes H as input and outputs $\hat{H} = \mathcal{M}(H)$, satisfying (ε, δ)-DP under the adopted adjacency notion. Our objective is to maximize the utility of $\hat{H}$ as an approximation to H while preserving spatial correlations and temporal continuity, which are important for analyses such as aggregation and pattern discovery. We evaluate the proposed mechanism on several representative tasks using the utility metrics in Section 7.

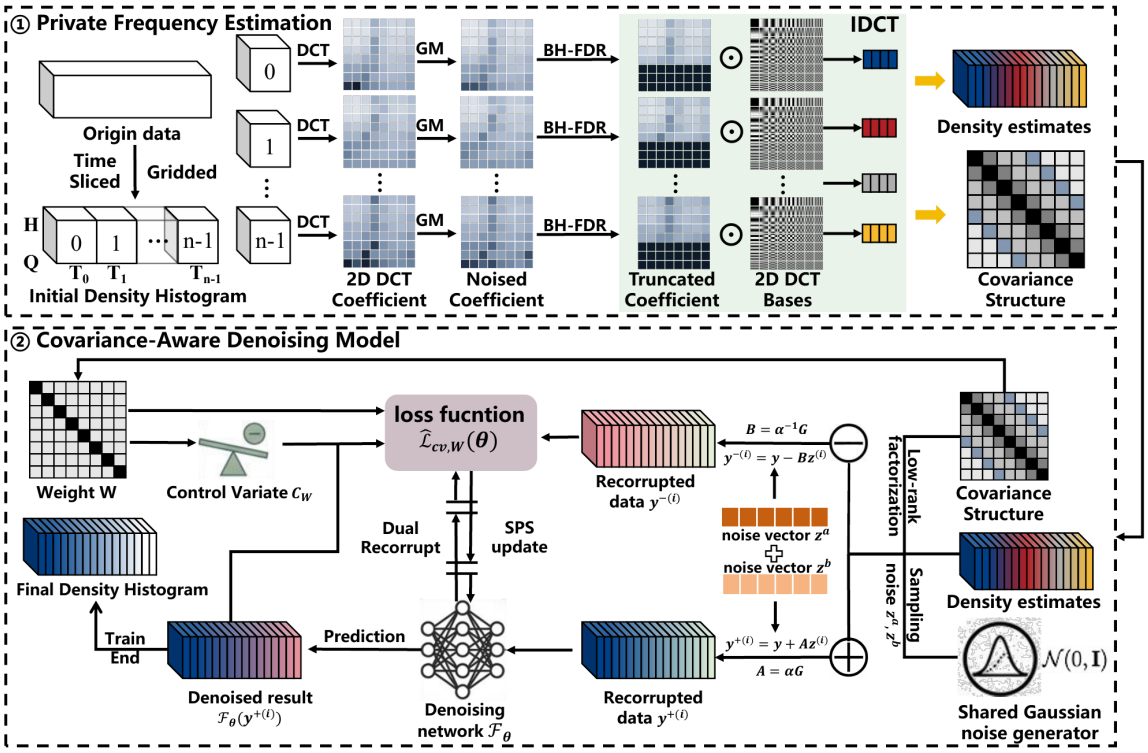

Figure 2: Framework of PrivSTD.

4 OVERVIEW OF PRIVSTD

As illustrated in Figure 2, PrivSTD is designed for the differentially private release of high-resolution spatiotemporal trajectory density data. It consists of two key components: *Private Frequency Estimation* and *Covariance-aware Denoising Model*.

Private Frequency Estimation. PrivSTD first discretizes trajectories into a spatiotemporal density histogram. For each time slice T_i, it applies a two-dimensional Discrete Cosine Transform (DCT) to map spatial densities into the frequency domain, where low-frequency coefficients capture dominant spatial correlations. The Gaussian mechanism (GM) perturbs the frequency coefficients to ensure (ε, δ)-DP, followed by a Benjamini–Hochberg false discovery rate (BH-FDR)-based truncation strategy to remove noise-dominated components. An inverse DCT (IDCT) then reconstructs preliminary private density estimates, while the retained DCT bases induce a covariance structure for the subsequent denoising stage.

Covariance-aware Denoising Model. Although the frequency-domain mechanism preserves dominant spatiotemporal structure, the IDCT-reconstructed estimates still contain structured and heteroscedastic DP noise. PrivSTD therefore factorizes the covariance structure to obtain a covariance factor G, samples two independent Gaussian noise vectors z^a and z^b, and combines them with a scaling parameter α to construct a statistically consistent pair of independent inputs, y^+ and y^-. The denoising network takes y^+ as input and uses y^- as a self-supervised target, forming a recorrupted-to-recorrupted objective between $\mathcal{F}_\theta(y^+)$ and y^-.

To improve robustness under heteroscedastic noise, PrivSTD further incorporates a covariance-induced precision weight matrix W and a control variate C_W, yielding the control-variate-augmented loss $\hat{\mathcal{L}}_{cv,W}(\theta)$. The dual-sample design also enables online estimation of stochastic gradient variability, allowing PrivSTD to perform *Stochastic Polyak Step-size* (SPS) updates that adapt the learning rate according to both loss suboptimality and noise-induced uncertainty.

5 PRIVATE FREQUENCY ESTIMATION

Private frequency estimation releases dominant spectral components of spatiotemporal density grids under differential privacy. Naive cell-wise perturbation ignores spatial smoothness, disperses variance across frequencies, and degrades utility [17, 40]. We therefore estimate spectra directly using a *Discrete Cosine Transform Mechanism (DCTM)*, which perturbs decorrelated coefficients with bounded ℓ_2-sensitivity and reconstructs sanitized densities through the inverse transform.

5.1 The Discrete Cosine Transform

The Discrete Cosine Transform (DCT) is a linear, invertible orthogonal transform with strong energy compaction [3, 37]. We use the DCT-II basis to represent grid cells as sums of cosine modes at increasing spatial frequencies. For a one-dimensional signal of length N, the DCT-II is

$$Y_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right], \quad k = 0, \ldots, N-1,$$

with standard orthonormal scaling.

For a grid count matrix $X \in \mathbb{R}^{H\times Q}$, the 2D DCT-II is computed separably:

$$\Theta_{u,v} = \sum_{i=0}^{H-1}\sum_{j=0}^{Q-1} X_{i,j}\, C_x[i,u]\, C_y[j,v], \quad u = 0,\dots,H-1,\ v = 0,\dots,Q-1, \tag{1}$$

or equivalently,

$$\Theta = C_x^\top X C_y,$$

where $C_x \in \mathbb{R}^{H\times H}$ and $C_y \in \mathbb{R}^{Q\times Q}$ are orthogonal cosine bases:

$$C_x[i,u] = \begin{cases} \frac{1}{\sqrt{H}}, & u = 0 \\ \sqrt{\frac{2}{H}}\cos\Big(\frac{\pi(2i+1)u}{2H}\Big), & u \geq 1 \end{cases},$$

$$C_y[j,v] = \begin{cases} \frac{1}{\sqrt{Q}}, & v = 0 \\ \sqrt{\frac{2}{Q}}\cos\Big(\frac{\pi(2j+1)v}{2Q}\Big), & v \geq 1 \end{cases}.$$

Orthogonality ($C_x^\top C_x = I_H$, $C_y^\top C_y = I_Q$) implies Parseval preservation, $\|X\|_F^2 = \|\Theta\|_F^2$, and concentrates energy in low-frequency coefficients near $\Theta_{0,0}$ [3]. This is well suited to trajectory densities with smooth spatial variation. Reconstruction uses inverse DCT:

$$X = C_x \Theta C_y^\top.$$

5.2 DCT under Differential Privacy

As a concrete instantiation of private frequency estimation on grid-structured data, DCTM perturbs DCT coefficients while bounding individual contributions. We present the derivation under event-level adjacency and then state the corresponding user-level extension under per-user contribution bounding.

Let $f(D) = \mathrm{vec}(\Theta)$ denote the query that outputs the flattened retained DCT coefficients, where D is the dataset of location reports aggregated into X. Under event-level adjacency, neighboring datasets $D \sim D'$ differ by the addition or removal of a single report, which changes exactly one grid cell by ± 1.

For retained bandwidths $U \leq H$ and $V \leq Q$, define

$$s_x(i;U) = \sum_{u<U} C_x[i,u]^2,$$
$$s_y(j;V) = \sum_{v<V} C_y[j,v]^2,$$
$$w_{ij}(U,V) = s_x(i;U)\, s_y(j;V).$$

Lemma 1. *For all i, j, U, V, $0 \leq s_x(i;U), s_y(j;V) \leq 1$. These quantities are nondecreasing in U and V, attain 1 at $U = H$ and $V = Q$, and satisfy $\sum_i s_x(i;U) = U$ and $\sum_j s_y(j;V) = V$.*

Proof. Since $C_x^\top C_x = I_H$, we also have $C_x C_x^\top = I_H$, so $\sum_u C_x[i,u]^2 = 1$ for each i. Hence $s_x(i;U)$, as a partial sum of nonnegative terms, lies in $[0,1]$, is nondecreasing in U, and satisfies $s_x(i;H) = 1$. The same argument applies to $s_y(j;V)$.

Moreover,

$$\sum_i s_x(i;U) = \sum_i \sum_{u<U} C_x[i,u]^2 = \sum_{u<U}\sum_i C_x[i,u]^2 = \sum_{u<U} 1 = U,$$

since each column of C_x has unit norm. Likewise, $\sum_j s_y(j;V) = V$. □

Theorem 1. *Under event-level adjacency, the global DCT ℓ_2-sensitivity is computed as:*

$$\Delta_2^{\text{event}}(U,V) = \sqrt{\max_i s_x(i;U) \cdot \max_j s_y(j;V)} \leq 1.$$

Proof. Let $X^{(t)} \in \mathbb{R}^{H\times Q}$ be the count matrix at time t, and $x^{(t)} = \mathrm{vec}(X^{(t)})$. The orthonormal 2D DCT of frame t is

$$\Theta^{(t)} = C_x^\top X^{(t)} C_y.$$

Retaining the $U \times V$ low-frequency coefficients gives

$$g^{(t)}(D) = \mathrm{ved}\big(\Theta^{(t)}_{0:U,\,0:V}\big) = \big(C_y^{(:,\,0:V)} \otimes C_x^{(:,\,0:U)}\big)^\top x^{(t)}.$$

For neighboring datasets D and D', there exists a unique (i_0, j_0, t_0) such that

$$x^{(t_0)} - x'^{(t_0)} = \pm e_{i_0 j_0}, \qquad x^{(t)} = x'^{(t)} \text{ for } t \neq t_0.$$

Hence only frame t_0 changes, and for $0 \leq u < U,\ 0 \leq v < V$,

$$\Theta^{(t_0)}_{u,v}(D) - \Theta^{(t_0)}_{u,v}(D') = \pm C_x[i_0,u]\, C_y[j_0,v].$$

Therefore,

$$\begin{aligned}\|f(D) - f(D')\|_2^2 &= \|g^{(t_0)}(D) - g^{(t_0)}(D')\|_2^2 \\ &= \sum_{u=0}^{U-1}\sum_{v=0}^{V-1} \big(C_x[i_0,u]\, C_y[j_0,v]\big)^2 \\ &= s_x(i_0;U)\, s_y(j_0;V).\end{aligned}$$

Taking the maximum over (i_0, j_0) yields

$$\Delta_2^{\text{event}}(U,V) = \sqrt{\max_i s_x(i;U) \cdot \max_j s_y(j;V)}.$$

The bound $\Delta_2^{\text{event}}(U,V) \leq 1$ follows from Lemma 1. □

Corollary 2 (User-level extension). *If each user contributes at most k reports, then under user-level adjacency, we can get*

$$\Delta_2^{\text{user}}(U,V;k) = k\,\Delta_2^{\text{event}}(U,V) = k\sqrt{\max_i s_x(i;U) \cdot \max_j s_y(j;V)} \leq k.$$

Proof. Under user-level adjacency, neighboring datasets differ by all reports of one user. If each user contributes at most k reports, the query difference is the sum of at most k event-level perturbations. The claim then follows from linearity and the triangle inequality. □

DCTM perturbs the retained DCT coefficients with i.i.d. Gaussian noise and reconstructs the private density estimate by inverse DCT. For (ε,δ)-DP, the Gaussian scale must satisfy

$$\sigma \geq \frac{\Delta_2\sqrt{2\ln(1.25/\delta)}}{\varepsilon}. \tag{2}$$

Under event-level adjacency, Theorem 1 gives $\Delta_2 \leq 1$, yielding $\sigma \geq \sqrt{2\ln(1.25/\delta)}/\varepsilon$; under user-level adjacency with at most k reports per user, Corollary 2 gives $\Delta_2 \leq k$, yielding $\sigma \geq k\sqrt{2\ln(1.25/\delta)}/\varepsilon$.

Theorem 3. *The Discrete Cosine Transform Mechanism satisfies (ε,δ)-DP.*

Proof. DCTM computes $f(D)$ and adds Gaussian noise with variance σ^2 calibrated as in Equation (2). For neighboring datasets D and D', the outputs are multivariate Gaussians with means $f(D)$ and $f(D')$ and common covariance $\sigma^2 I$. The conclusion then follows from the standard Gaussian mechanism analysis for ℓ_2-sensitivity [19], using Theorem 1 for event-level adjacency and Corollary 2 for user-level adjacency. □

The inverse DCT is post-processing and incurs no additional privacy loss [19]. By orthogonality, noise energy is preserved under the transform; equivalently, reconstruction noise remains Gaussian with variance σ^2 per cell up to an orthogonal mixing.

5.3 BH-FDR Truncation

DP injects Gaussian noise across the spectrum, so high-frequency coefficients are often noise-dominated. Because DCT bases have global support, this noise propagates across the spatial domain after inversion [31]. Frequency-domain truncation mitigates this effect, but fixed bandwidths do not adapt to dataset spectra or privacy noise levels [15]. We therefore apply BH-FDR as a post-processing step on DP spectra to adaptively select an effective bandwidth without additional privacy cost.

The false discovery rate (FDR) is

$$\mathrm{FDR} = \mathbb{E}\left[\frac{V}{\max(R,1)}\right],$$

where V is the number of false rejections and R is the total number of rejections [5]. We use BH-FDR because it is less conservative than FWER while still controlling the expected false-discovery proportion in high dimensions [39]. The BH procedure sorts p-values $p_{(1)} \le \cdots \le p_{(m)}$ and selects the largest k such that $p_{(k)} \le (k/m)q$, rejecting all $p \le p_{(k)}$; it controls FDR under independence and admits extensions to correlated tests [6].

In our DP-DCT framework,

$$\Theta^{\mathrm{full}}_{\mathrm{noisy}} = \Theta_{\mathrm{true}} + Z, \qquad Z_{uv} \sim \mathcal{N}(0,\sigma^2), \qquad \sigma = \frac{\Delta_2\sqrt{2\ln(1.25/\delta)}}{\epsilon}.$$

To exploit temporal redundancy across T frames, we define

$$S(u,v) = \sum_{t=1}^{T} \frac{\left(\Theta^{(t)}_{\mathrm{noisy}}[u,v]\right)^2}{\sigma^2}.$$

Under the null that only DP noise is present, $S(u,v) \sim \chi^2_T$, yielding

$$p(u,v) = 1 - F_{\chi^2_T}(S(u,v)),$$

where $F_{\chi^2_T}$ is the chi-squared CDF [7]. Applying BH to the $m = HQ$ p-values yields a set $\mathcal{S}$ of frequency coordinates whose energy is unlikely to be explained by DP noise alone.

Figure 3 illustrates this process. Figure 3(a) shows the noisy 2D DCT coefficients, where $(0,0)$ is the DC component and spatial frequency increases toward the bottom-right corner. Figure 3(b) highlights the coordinates selected by BH-FDR, and their outermost positions determine the effective frequency band. Figure 3(c) shows the truncated coefficients after discarding entries outside the retained band. This suppresses high-frequency noise and reduces its propagation after inverse DCT.

Rather than retaining only the pointwise significant set $\mathcal{S}$, we adopt a rectangularized low-pass region to avoid fragmented masks

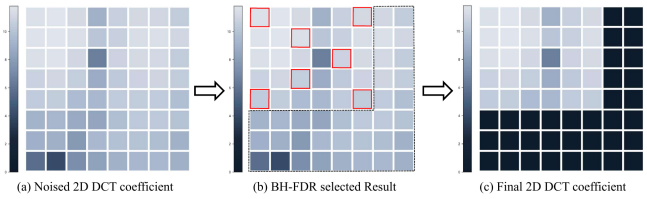

Figure 3: BH-FDR Example.

that create spatial artifacts. Under low-frequency dominance and spectral connectivity [34], if $(u,v) \in \mathcal{S}$ with large u or v, then lower-frequency coefficients typically also carry informative structure, even if some fail the BH test because of DP noise. We define

$$U^* = \max_{(u,v)\in\mathcal{S}} u, \qquad V^* = \max_{(u,v)\in\mathcal{S}} v,$$

and retain the region $[0:U^*] \times [0:V^*]$ as the final bandwidth. The truncated coefficients are

$$\Theta^{\mathrm{trunc}}_{\mathrm{noisy}}[u,v] = \begin{cases} \Theta^{\mathrm{full}}_{\mathrm{noisy}}[u,v], & (u,v) \in [0:U^*] \times [0:V^*], \\ 0, & \text{otherwise}, \end{cases}$$

with reconstruction

$$\widetilde{X} = C_x \Theta^{\mathrm{trunc}}_{\mathrm{noisy}} C_y^{\top}.$$

By Parseval's identity [37], the expected post-truncation noise energy is $\mathbb{E}[\|\widetilde{E}\|_F^2] = U^*V^*\sigma^2$, compared with $HQ\sigma^2$ without truncation. At the same time, FDR control helps preserve coefficients carrying genuine signal energy and mitigates over-aggressive truncation bias. For dependent tests, we use the Benjamini–Yekutieli adjustment $q \leftarrow q/\sum_{i=1}^{m} 1/i$ to control FDR under arbitrary dependence [6].

Algorithm 1 implements private frequency estimation with DCT perturbation and BH-FDR truncation. It first computes the Gaussian noise scale σ and, for each time frame, applies the 2D DCT and adds i.i.d. Gaussian noise to obtain noisy coefficients $\Theta^{(t)}_{\mathrm{noisy}}$ (lines 1–4). It then computes the energy statistic $S(u,v)$ and the corresponding p-value for each frequency coordinate, sorts all p-values, and applies the BH procedure to identify the significant set $\mathcal{S}$ (lines 5–9). The retained bandwidths U^* and V^* are obtained by rectangularizing $\mathcal{S}$ (lines 10–11). Finally, for each time frame, the algorithm keeps only the coefficients within $[1:U^*, 1:V^*]$, sets all others to zero, and returns the truncated noisy coefficients (lines 12–14).

6 COVARIANCE-AWARE DENOISING MODEL

Differential privacy inevitably degrades utility through injected noise. In our setting, the perturbation entering denoising is structured, heteroscedastic, and further shaped by the adaptive truncation mechanism of Section 5. Because the retained frequency region is selected from noisy DCT coefficients via BH-FDR and rectangularization, the resulting reconstruction is governed by a data-dependent mask rather than a fixed Gaussian covariance model. This prevents classical exact R2R supervised equivalence from applying directly. To address this, we develop a covariance-aware self-supervised denoising framework with three components: a control-variate estimator for variance reduction, a weighted GLS objective for heterogeneous uncertainty together with an approximate supervised alignment theorem, and a variance-aware Polyak

Algorithm 1: Frequency Noised Mechanism

Input: Grid matrices $\{X^{(t)}\}_{t=1}^{T} \in \mathbb{R}^{H\times Q}$, privacy parameters (ϵ, δ), FDR level q, DCT matrices C_x, C_y

Output: Truncated noisy coefficients $\{\Theta_{\text{noisy}}^{(t),\text{trunc}}\}_{t=1}^{T}$

$\sigma \leftarrow \dfrac{\Delta_2\sqrt{2\ln(1.25/\delta)}}{\epsilon}$;

for $t \leftarrow 1$ **to** T **do**

  $\Theta_{\text{true}}^{(t)} \leftarrow C_x^{\top} X^{(t)} C_y$; sample $Z^{(t)} \sim \mathcal{N}(0,\sigma^2)^{H\times Q}$;

  $\Theta_{\text{noisy}}^{(t)} \leftarrow \Theta_{\text{true}}^{(t)} + Z^{(t)}$;

for $u \leftarrow 1$ **to** H **do**

  for $v \leftarrow 1$ **to** Q **do**

    $S(u,v) \leftarrow \sum_{t=1}^{T} \left(\Theta_{\text{noisy}}^{(t)}[u,v]\right)^2/\sigma^2$;

    $p(u,v) \leftarrow 1 - F_{\chi^2_T}(S(u,v))$;

Set $m \leftarrow HQ$ and sort $\{p(u,v)\}$ as $p_{(1)} \le \cdots \le p_{(m)}$;

Find $k^\star \leftarrow \max\{k \mid p_{(k)} \le (k/m)\, q\}$;

$\mathcal{S} \leftarrow \{(u,v) \mid p(u,v) \le p_{(k^\star)}\}$;

$U^* \leftarrow \max_{(u,v)\in\mathcal{S}} u$; $\quad V^* \leftarrow \max_{(u,v)\in\mathcal{S}} v$;

for $t \leftarrow 1$ **to** T **do**

  $\Theta_{\text{noisy}}^{(t),\text{trunc}} \leftarrow 0_{H\times Q}$;

  $\Theta_{\text{noisy}}^{(t),\text{trunc}}[1:U^*, 1:V^*] \leftarrow \Theta_{\text{noisy}}^{(t)}[1:U^*, 1:V^*]$;

return $\{\Theta_{\text{noisy}}^{(t),\text{trunc}}\}_{t=1}^{T}$;

step size based on dual recorruption samples. Unless otherwise stated, symbols with subscript W denote the weighted versions of their unweighted counterparts.

6.1 Variance Reduction via Control Variates

A central difficulty in self-supervised denoising is stochastic instability: each recorruption resamples noise and injects variance into both the loss and its gradient. In our setting, however, the retained DCT structure remains available from private frequency estimation, enabling an analytic control variate that is centered, inexpensive to evaluate, and strongly correlated with the stochastic loss. Let $\mathcal{M}$ denote the random retained mask induced by BH-FDR and rectangularization, and write the resulting private observation as $y = x + n_{\mathcal{M}}$. In the denoising stage, we condition on the realized retained region $[0{:}U)\times[0{:}V)$; under this conditioning, the matrices below are fixed. Define

$$G = \sigma\big(C_y^{(:,0:V)} \otimes C_x^{(:,0:U)}\big) \in \mathbb{R}^{d\times K}, \qquad K = UV,$$

and let

$$z \sim \mathcal{N}(0, I_K).$$

For $\alpha > 0$, set

$$A = \alpha G, \qquad B = \alpha^{-1} G.$$

The recorrupted pair is

$$y^+ = y + Az, \qquad y^- = y - Bz, \tag{3}$$

and the adopted self-supervised loss is

$$\mathcal{L}(\theta) = \|\mathcal{F}_\theta(y^+) - y^-\|_2^2. \tag{4}$$

Using $\Delta = y^+ - y^-$, we have

$$\Delta = (A+B)z = G\gamma z, \qquad \gamma = \alpha + \alpha^{-1}.$$

This motivates the quadratic control variate

$$C = \|\Delta\|_2^2 = \gamma^2 z^{\top}(G^{\top}G)z.$$

Since $\mathcal{L}(\theta)$ and C share the same recorruption realization, they are typically correlated. We therefore consider the centered estimator

$$\widehat{\mathcal{L}}_{\text{cv}}(\theta) = \mathcal{L}(\theta) - \lambda\big(C - \mathbb{E}[C]\big).$$

The following theorem shows that this correction preserves the expectation of the adopted self-supervised objective while achieving optimal variance reduction within the class of linear control variates.

Theorem 4 (Unbiasedness and optimal variance reduction of the CV estimator). *Assume that $\mathcal{L}(\theta)$ and C have finite second moments. Then*

$$\mathbb{E}[\widehat{\mathcal{L}}_{\text{cv}}(\theta)] = \mathbb{E}[\mathcal{L}(\theta)].$$

Moreover, the coefficient minimizing $\operatorname{Var}(\widehat{\mathcal{L}}_{\text{cv}})$ is

$$\lambda^\star = \frac{\operatorname{Cov}(\mathcal{L}, C)}{\operatorname{Var}(C)},$$

and the resulting minimum variance is

$$\operatorname{Var}(\widehat{\mathcal{L}}_{\text{cv}}) = \operatorname{Var}(\mathcal{L})(1-\rho^2_{\mathcal{L},C}), \qquad \rho_{\mathcal{L},C} = \frac{\operatorname{Cov}(\mathcal{L}, C)}{\sqrt{\operatorname{Var}(\mathcal{L})\operatorname{Var}(C)}}. \tag{5}$$

If $z \sim \mathcal{N}(0, I_K)$, then

$$\mathbb{E}[C] = \gamma^2 \operatorname{tr}(G^{\top}G), \qquad \operatorname{Var}(C) = 2\gamma^4 \operatorname{tr}\big((G^{\top}G)^2\big).$$

Under the present construction of G,

$$G^{\top}G = \sigma^2 I_K, \qquad \operatorname{Var}(C) = 2\gamma^4\sigma^4 K > 0.$$

Proof. Since $C - \mathbb{E}[C]$ has zero expectation,

$$\mathbb{E}[\widehat{\mathcal{L}}_{\text{cv}}(\theta)] = \mathbb{E}[\mathcal{L}(\theta)].$$

Also,

$$\operatorname{Var}(\widehat{\mathcal{L}}_{\text{cv}}) = \operatorname{Var}(\mathcal{L}) + \lambda^2 \operatorname{Var}(C) - 2\lambda \operatorname{Cov}(\mathcal{L}, C).$$

Minimizing this quadratic in λ gives

$$\lambda^\star = \frac{\operatorname{Cov}(\mathcal{L}, C)}{\operatorname{Var}(C)},$$

and substitution yields

$$\operatorname{Var}(\widehat{\mathcal{L}}_{\text{cv}}) = \operatorname{Var}(\mathcal{L}) - \frac{\operatorname{Cov}^2(\mathcal{L}, C)}{\operatorname{Var}(C)} = \operatorname{Var}(\mathcal{L})(1-\rho^2_{\mathcal{L},C}).$$

Let $S = G^{\top}G \succeq 0$. Since $C = \gamma^2 z^{\top} S z$, Gaussian quadratic-form identities give

$$\mathbb{E}[z^{\top}Sz] = \operatorname{tr}(S), \qquad \operatorname{Var}(z^{\top}Sz) = 2\operatorname{tr}(S^2),$$

hence

$$\mathbb{E}[C] = \gamma^2 \operatorname{tr}(G^{\top}G), \qquad \operatorname{Var}(C) = 2\gamma^4 \operatorname{tr}\big((G^{\top}G)^2\big).$$

Finally, since $C_x^{(:,0:U)}$ and $C_y^{(:,0:V)}$ have orthonormal columns,

$$G^{\top}G = \sigma^2\Big((C_y^{(:,0:V)})^{\top} C_y^{(:,0:V)}\Big) \otimes \Big((C_x^{(:,0:U)})^{\top} C_x^{(:,0:U)}\Big) = \sigma^2 I_K,$$

so $\operatorname{Var}(C) = 2\gamma^4\sigma^4 K > 0$. □

Corollary 5 (Scaling that minimizes the quadratic-form variance factor). *For $\alpha \in \mathbb{R}_{>0}$, the factor*

$$\gamma^4 = (\alpha + \alpha^{-1})^4$$

is uniquely minimized at $\alpha = 1$.

Proof. Let $g(\alpha) = \alpha + \alpha^{-1}$ for $\alpha > 0$. Then

$$g'(\alpha) = 1 - \alpha^{-2}, \qquad g''(\alpha) = 2\alpha^{-3} > 0.$$

Hence g is strictly convex and has a unique stationary point at $\alpha = 1$, which is its unique global minimizer. Since $u \mapsto u^4$ is strictly increasing on $\mathbb{R}_{>0}$, the same holds for $\gamma^4 = g(\alpha)^4$. □

6.2 Approximate GLS Alignment

Adaptive truncation induces heterogeneous uncertainty across spatial cells, since the effective reconstruction variance depends on the retained basis structure. We therefore extend the self-supervised objective to a weighted generalized least squares (GLS) form, allowing the denoising objective to reflect this heteroscedasticity and to be related to the desired supervised risk in a bias-controlled manner.

Define

$$W := \mathrm{Diag}\left(\frac{1}{\sigma^2\, w_{ij}(U,V) + \varepsilon_w}\right)_{i,j}, \qquad w_{ij}(U,V) = s_x(i;U)s_y(j;V),$$

where $\varepsilon_w > 0$ is a numerical-stability parameter. Then W is fixed and symmetric positive semidefinite under the realized retained region, and we define

$$\mathcal{L}_W(\theta) = \|W^{1/2}(\mathcal{F}_\theta(y^+) - y^-)\|_2^2. \tag{6}$$

We also define the weighted control variate

$$C_W = \Delta^\top W \Delta = \gamma^2 z^\top (G^\top W G) z,$$

and the weighted CV estimator

$$\widehat{\mathcal{L}}_{\mathrm{cv},W}(\theta) = \mathcal{L}_W(\theta) - \lambda\big(C_W - \mathbb{E}[C_W]\big),$$

with

$$\mathbb{E}[C_W] = \gamma^2 \,\mathrm{tr}(G^\top W G).$$

To relate the weighted self-supervised objective to the desired supervised risk, define

$$R_W(\theta) = \mathbb{E}\|W^{1/2}(\mathcal{F}_\theta(y^+) - x)\|_2^2, \qquad K_W = \mathbb{E}\|W^{1/2}(x - y^-)\|_2^2,$$

and let

$$b^\star(y^+, x) = \mathbb{E}[x - y^- \mid y^+, x], \qquad \delta_W^2 = \mathbb{E}\|W^{1/2} b^\star(y^+, x)\|_2^2.$$

The next theorem shows that the weighted self-supervised objective approximates the corresponding supervised GLS risk up to a bias term controlled by δ_W.

Theorem 6 (Approximate weighted GLS alignment). *Assume the expectations below are finite. Then*

$$\begin{aligned}\mathbb{E}[\mathcal{L}_W(\theta)] &= R_W(\theta) + K_W \\ &\quad + 2\,\mathbb{E}\Big[\big\langle W^{1/2}(\mathcal{F}_\theta(y^+) - x), W^{1/2} b^\star(y^+, x)\big\rangle\Big].\end{aligned} \tag{7}$$

Consequently,

$$|\mathbb{E}[\mathcal{L}_W(\theta)] - R_W(\theta) - K_W| \le 2\sqrt{R_W(\theta)}\,\delta_W. \tag{8}$$

Moreover, for any $\eta > 0$,

$$\mathbb{E}[\mathcal{L}_W(\theta)] \le (1+\eta) R_W(\theta) + K_W + \eta^{-1}\delta_W^2, \tag{9}$$

and

$$\mathbb{E}[\mathcal{L}_W(\theta)] \ge (1-\eta) R_W(\theta) + K_W - \eta^{-1}\delta_W^2. \tag{10}$$

Hence, when δ_W is small, the weighted self-supervised objective is a controlled approximation to the weighted supervised GLS risk up to the additive constant K_W.

Proof. Let

$$a = \mathcal{F}_\theta(y^+) - x, \qquad b = x - y^-.$$

Then

$$\mathcal{F}_\theta(y^+) - y^- = a + b,$$

so

$$\mathcal{L}_W(\theta) = \|W^{1/2}(a+b)\|_2^2 = \|W^{1/2}a\|_2^2 + \|W^{1/2}b\|_2^2 + 2\langle W^{1/2}a, W^{1/2}b\rangle.$$

Taking expectations gives

$$\mathbb{E}[\mathcal{L}_W(\theta)] = R_W(\theta) + K_W + 2\mathbb{E}\langle W^{1/2}a, W^{1/2}b\rangle.$$

Since a is measurable with respect to (y^+, x), the tower property yields

$$\mathbb{E}\langle W^{1/2}a, W^{1/2}b\rangle = \mathbb{E}\Big[\big\langle W^{1/2}(\mathcal{F}_\theta(y^+) - x), W^{1/2} b^\star(y^+, x)\big\rangle\Big],$$

proving the decomposition.

Applying Cauchy–Schwarz,

$$|\mathbb{E}[\mathcal{L}_W(\theta)] - R_W(\theta) - K_W| \le 2\sqrt{R_W(\theta)}\,\delta_W.$$

The upper and lower bounds then follow from Young's inequality

$$2ab \le \eta a^2 + \eta^{-1} b^2, \qquad \eta > 0,$$

with $a = \sqrt{R_W(\theta)}$ and $b = \delta_W$. □

Remark. For the special case $y = x + n$, exact alignment holds in expectation; under adaptive truncation, δ_W quantifies the deviation from exact alignment.

6.3 Variance-Aware Polyak Step Size

Although the control-variate estimator reduces the variance of the stochastic objective, optimization remains difficult under heterogeneous privacy noise and evolving signal-to-noise ratios. Dual recorruption naturally provides two independent gradient realizations, which together yield both an unbiased gradient estimator and an online proxy for gradient-noise intensity. This motivates a variance-aware Polyak rule.

6.3.1 Dual Recorruption Architecture and Variance Estimation. At iteration t, sample two independent auxiliary noises

$$z_t^{(a)}, z_t^{(b)} \sim \mathcal{N}(0, I_K), \qquad z_t^{(a)} \perp z_t^{(b)}.$$

These generate two recorrupted pairs

$$y_t^{+(i)} = y + A z_t^{(i)}, \qquad y_t^{-(i)} = y - B z_t^{(i)}, \qquad i \in \{a, b\}.$$

Let $\ell^{(i)}(\theta)$ denote the stochastic loss on the i-th pair; this may be $\mathcal{L}$, $\mathcal{L}_W$, $\widehat{\mathcal{L}}_{\mathrm{cv}}$, or $\widehat{\mathcal{L}}_{\mathrm{cv},W}$. Define

$$g_t^{(i)} = \nabla_\theta \ell^{(i)}(\theta_t), \qquad \bar{g}_t = \frac{1}{2}(g_t^{(a)} + g_t^{(b)}), \qquad v_t = \frac{1}{4}\|g_t^{(a)} - g_t^{(b)}\|_2^2.$$

The following theorem shows that the two-sample construction provides both an unbiased gradient estimator and a simple variance proxy.

Theorem 7 (Unbiased gradient and a variance proxy). *Assume that, conditional on (θ_t, y), the only randomness in $g_t^{(a)}$ and $g_t^{(b)}$ comes from the independent auxiliary noises $z_t^{(a)}$ and $z_t^{(b)}$, and that the required second moments are finite. Then*

$$\mathbb{E}[\bar{g}_t \mid \theta_t, y] = \mathbb{E}[g_t \mid \theta_t, y],$$

and

$$\mathbb{E}[v_t \mid \theta_t, y] = \frac{1}{2}\mathbb{E}\big[\|g_t - \mathbb{E}[g_t \mid \theta_t, y]\|_2^2 \,\big|\, \theta_t, y\big].$$

Thus $\bar{g}_t$ is an unbiased gradient estimator for the adopted stochastic objective, while v_t is an unbiased proxy for the conditional gradient variance up to the factor $1/2$.

Proof. The first claim follows from linearity of expectation and the fact that the two gradients are conditionally identically distributed. The second claim follows by expanding $\|g_t^{(a)} - g_t^{(b)}\|_2^2$ and using the conditional independence of the two gradient realizations. A full proof is available in the full version [46]. □

6.3.2 Variance-Aware Stochastic Polyak Rule. Motivated by the preceding theorem, we use the practical step-size rule

$$\eta_t = \min\left(\frac{\hat{\mathcal{L}}(\theta_t) - L_{\text{lb}}}{\|\bar{g}_t\|_2^2 + \beta v_t + \varepsilon}, \eta_{\max}\right),$$

where $\hat{\mathcal{L}}(\theta_t) = \frac{1}{2}(\ell^{(a)}(\theta_t) + \ell^{(b)}(\theta_t))$ is the two-sample loss estimate, L_{lb} is a user-specified lower bound or lower-bound surrogate, $\beta > 0$ controls conservatism with respect to gradient noise, $\varepsilon > 0$ ensures numerical stability, and $\eta_{\max} > 0$ caps the step size in poorly conditioned regions.

The next proposition provides a descent guarantee for the corresponding idealized population-level rule.

Proposition 1 (Expected descent of an idealized variance-regularized Polyak step). *Assume that $J : \mathbb{R}^p \to \mathbb{R}$ is L_J-smooth and satisfies the Polyak–Łojasiewicz condition*

$$\frac{1}{2}\|\nabla J(\theta)\|_2^2 \geq \mu\big(J(\theta) - J_\star\big), \qquad \mu > 0.$$

Assume also that the stochastic gradient satisfies

$$\mathbb{E}[g_t \mid \theta_t] = \nabla J(\theta_t), \qquad \mathbb{E}\|g_t - \nabla J(\theta_t)\|_2^2 \mid \theta_t \leq \sigma^2.$$

Consider the idealized clamped step size

$$\eta_t = \min\left(\frac{J(\theta_t) - J_\star}{\|\nabla J(\theta_t)\|_2^2 + \beta\sigma^2}, \frac{1}{L_J}\right), \qquad \beta \geq \frac{L_J}{\mu}.$$

Then the update $\theta_{t+1} = \theta_t - \eta_t g_t$ satisfies

$$\mathbb{E}[J(\theta_{t+1}) - J_\star \mid \theta_t] \leq \left(1 - \frac{\mu}{2}\eta_t\right)\big(J(\theta_t) - J_\star\big).$$

Proof. The result follows from a standard smoothness-based descent argument, combined with the unbiasedness and bounded variance of the stochastic gradient, the definition of the variance-regularized Polyak step size, and the Polyak–Łojasiewicz inequality. A full proof is available in the full version [46]. □

Table 2: Summary of datasets used in the experiments

Datasets	#Type	#Area	#Scale	#Users
4SQ_TKY	Event	Tokyo	755,885	-
CABS_SF	Event	San Francisco	846,000	-
Gowalla_SF	Event	San Francisco	568,000	-
Bri_TKY	User	Tokyo	133,593	2,105
Stockholm	User	Stockholm	141,262	8,780
Austin	User	Austin	212,514	7,099

Table 3: Parameter table

Parameter	Range
the privacy budget ε	0.1, 0.5,1, 2, **3**, 4, 5
the grid number H	128, 256, **384**, 512, 640, 768
the per-user contribution bound k_s	3,9,15,**21**

Practical implementation. In practice, we replace the population quantities $J(\theta_t)$ and $\nabla J(\theta_t)$ by their two-sample estimators $\hat{\mathcal{L}}(\theta_t)$ and $\bar{g}_t$, and replace the unknown conditional gradient variance by the proxy v_t. The theoretical cap $1/L_J$ is implemented as the hyperparameter $\eta_{\max}$, with a small $\varepsilon > 0$ added for numerical stability. This yields a robust variance-aware step-size rule that remains consistent with the idealized descent analysis while being fully implementable from stochastic dual-recorruption samples.

7 EXPERIMENTS

We evaluate PrivSTD against seven baselines on three representative tasks: **Cell Count**, **Hotspot Query**, and **Forecasting Query**. Our experiments cover both **event-level** and **user-level** privacy settings. To keep the presentation clear, we report the two settings in separate result tables, since they correspond to different privacy units and calibration protocols. Due to space constraints, additional ablation results are deferred to the appendix of the full version [46].

7.1 Experimental Setup

Datasets. Table 2 summarizes the datasets used in our experiments. For the event-level setting, we use three public real-world spatiotemporal datasets: **4SQ_TKY** [45], the Tokyo subset of the Foursquare dataset, containing 755,885 location updates from 8,000 users over 22 months; **Gowalla_SF** [12], the San Francisco subset of the Gowalla dataset, containing 568,000 location updates from 14,000 users over 575 days; and **CABS_SF** [33], a taxi GPS trajectory dataset in San Francisco with 846,000 records over 575 days. For the user-level setting, we use **Bri_TKY** [12], the Tokyo subset of the Brightkite dataset, containing 133,593 check-in records from 2,105 users over 30 months; **Stockholm** [12], the Stockholm subset of the Gowalla dataset, containing 141,262 location updates from 8,780 users over 443 days; and **Austin** [12], the Austin subset of the Gowalla dataset, containing 212,514 location updates from 7,099 users over 589 days. To align the protocols across privacy notions, both settings follow the same spatial and temporal discretization as in VDR [4]: the spatial domain is partitioned into grids from 128×128 to 768×768, and the temporal domain is discretized into 24 slices. For the user-level setting, we further follow VDR by limiting each user to at most k_s reports.

Experimental Settings. Our method is implemented in Python 3.9, and the experiments are conducted on a machine with an Intel Xeon 2.1GHz CPU, 128 GB main memory, and an RTX3090 GPU. The source code is publicly available [46].

Evaluation Metrics. We evaluate the released spatiotemporal density tensor on three representative tasks: **Cell Count**, **Hotspot Query**, and **Temporal Forecasting**. Unless otherwise stated, all results are averaged over a large set of randomized queries.

Cell Count. For randomly sampled cells (i, j, t), we evaluate accuracy using **relative error (RE)**, defined as $\mathrm{RE}(y, u) = \frac{|y-u|}{\max\{u,\psi\}}$, where $u = H[i, j, t]$ is the true count, $y = \hat{H}[i, j, t]$ is the released count, and ψ is a small constant used to stabilize the metric for low-count cells. **Hotspot Query.** Given a query location q, a threshold v, and a bounded region SR, we return the nearest cell in SR whose density exceeds v (or the maximum-density cell if no such cell exists), and evaluate the result using **mean absolute error (MAE)** between the true and released hotspot distances. **Forecasting Query.** We forecast each cell's future count with horizon $h = 3$ under a holdout protocol and evaluate accuracy using **symmetric mean absolute percentage error (sMAPE)**, defined as $\mathrm{sMAPE} = \frac{1}{h}\sum_t \frac{|A_t - F_t|}{(|A_t|+|F_t|)/2}$, where A_t denotes the true count from H and F_t denotes the prediction based on the released histogram $\hat{H}$. With monthly aggregation ($T = 24$), we set the seasonal period to $sp = 3$ and keep only time series with length at least $3sp$ and significant autocorrelation at lag sp under a 90% confidence level.

Baseline Methods. We compare against seven state-of-the-art differentially private methods: (1) **Uniform Grid (UG)** [35]; (2) **Adaptive Grid (AG)** [35]; (3) **HB-Striped** [36]; (4) **PrivBayes** [49]; (5) **AHP** [50]; (6) **MWEM** [21]; and (7) **VDR** [4], which uses VQ-VAE with multi-resolution learning as a post-processing denoiser. For the **event-level** setting, we follow the protocol adopted in VDR [4]: each record is treated as if it were contributed by a distinct user, so that all methods are compared under the same event-level adjacency notion and calibrated with the Gaussian mechanism under the same privacy parameters. For the **user-level** setting, we also follow VDR [4]. Specifically, VDR is evaluated with its original user-level design, which combines its dedicated sampling/truncation procedure with its estimation strategy. The other methods, including PrivSTD, are evaluated under a standard bounded-contribution protocol, where each user contributes at most k_s reports before privacy calibration under the same user-level adjacency notion. This ensures a fair comparison under a common user-level privacy definition while preserving VDR's original design.

7.2 Overall Findings

Tables 4 and 5 report the overall utility of PrivSTD under event-level and user-level privacy, respectively. Overall, PrivSTD achieves the best utility across all three query tasks under event-level privacy and remains highly competitive under the stronger user-level.

Result Visualization. Figure 4 shows heatmaps of released density data on the CABS_SF dataset, where brighter colors indicate higher densities. PrivSTD recovers density patterns that are visibly closer to the ground truth than the baselines.

Results in terms of Cell Count. Under event-level privacy, PrivSTD consistently achieves the lowest relative error across datasets, especially under stronger privacy constraints. Grid-based methods such as UG, AG, and HB-Striped perturb counts directly in the spatial domain and process time slices largely independently, which disrupts spatiotemporal structure and amplifies error at high resolution. PrivBayes also struggles to capture complex dependencies.

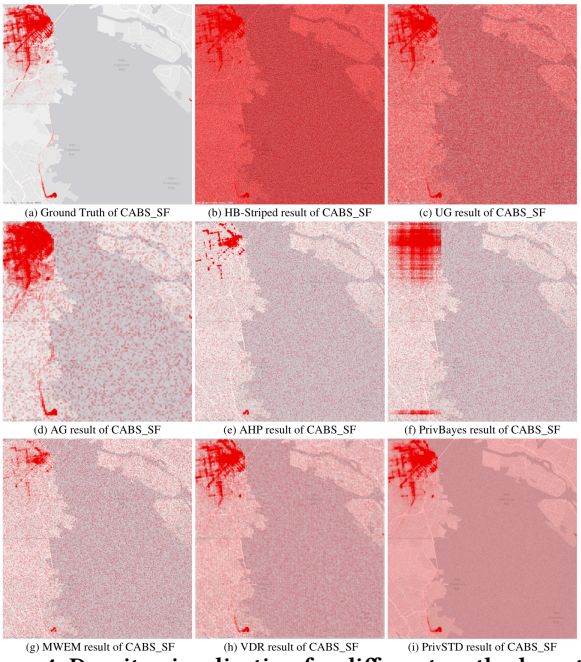

Figure 4: Density visualization for different methods on the CABS_SF dataset with $\epsilon = 0.5$.

Although VDR improves substantially over these baselines, it still injects noise in the spatial domain, so structural information is degraded before denoising. In contrast, PrivSTD preserves dominant low-frequency structure and then applies a covariance-aware denoising model to obtain more accurate density estimates. Under user-level privacy, Cell Count becomes more challenging because the stronger privacy notion increases sensitivity and lowers the signal-to-noise ratio. Since Cell Count is dominated by stable low-frequency density mass, PrivSTD benefits from spectral filtering and covariance-aware denoising, while spatial-domain methods suffer more from noisy cell-level perturbation.

Results in terms of Hotspot Query. PrivSTD also achieves the lowest hotspot localization error under event-level privacy. Spatial-domain perturbation in UG, AG, AHP, and MWEM tends to blur sharp density gradients and hotspot boundaries, especially in heterogeneous regions. VDR improves over these methods, but remains less effective at preserving local extrema after spatial-domain perturbation. In contrast, PrivSTD combines frequency-domain noise suppression with variance-reduced reconstruction, allowing it to preserve stronger local contrast and recover hotspot locations more accurately. Under user-level privacy, Hotspot Query is more sensitive to which structures survive perturbation. Although bounded contribution still trades completeness for noise, the main challenge is preserving localized peaks, sharp transitions, and neighborhood contrast. Because these signals partly reside in high-frequency DCT components, PrivSTD's truncation stabilizes the global density field but may attenuate peak-like evidence for hotspot localization. VDR can be competitive in some settings because learned spatial representations may preserve certain local patterns, but its bounded-sampling design does not guarantee consistent gains from more user reports. This suggests a future direction of combining global density reconstruction with privacy-aware local residual modeling.

Table 4: Overall utility under event-level privacy with different privacy budgets (best in bold; smaller is better).

Metric	Method	4SQ_TKY				CABS_SF				Gowalla_SF			
		$\epsilon = 0.5$	$\epsilon = 1.0$	$\epsilon = 2.0$	$\epsilon = 3.0$	$\epsilon = 0.5$	$\epsilon = 1.0$	$\epsilon = 2.0$	$\epsilon = 3.0$	$\epsilon = 0.5$	$\epsilon = 1.0$	$\epsilon = 2.0$	$\epsilon = 3.0$
Cell Count (RE)	UG	1.279	0.487	0.164	0.137	1.211	0.469	0.364	0.335	1.581	0.649	0.235	0.227
	AG	0.348	0.342	0.310	0.292	0.899	0.698	0.501	0.454	0.471	0.421	0.373	0.335
	HB-Striped	1.461	0.731	0.365	0.243	1.510	0.755	0.335	0.251	1.565	0.782	0.391	0.261
	AHP	0.407	0.363	0.292	0.248	0.513	0.507	0.446	0.369	0.409	0.404	0.373	0.321
	MWEM	0.385	0.369	0.342	0.314	0.473	0.454	0.420	0.390	0.382	0.369	0.342	0.319
	VDR	0.237	0.148	0.110	0.093	0.299	0.185	0.174	0.119	0.239	0.167	0.121	0.106
	PrivBayes	0.351	0.349	0.310	0.304	0.349	0.344	0.333	0.333	0.282	0.275	0.270	0.256
	PrivSTD	**0.149**	**0.119**	**0.073**	**0.057**	**0.280**	**0.154**	**0.099**	**0.075**	**0.194**	**0.128**	**0.086**	**0.067**
Hotspot Query (MAE)	UG	10.023	8.092	1.851	1.296	5.267	4.191	3.897	3.172	6.238	5.732	4.970	2.397
	AG	5.454	5.172	4.703	4.921	4.368	2.574	2.147	2.238	4.809	4.596	3.858	3.172
	HB-Striped	10.378	10.203	9.497	7.727	3.968	3.805	3.458	2.753	5.765	5.588	5.054	3.724
	AHP	25.223	11.142	4.578	1.841	42.258	14.120	2.817	0.921	95.914	23.819	6.014	0.811
	MWEM	21.553	19.611	12.114	11.207	28.083	24.789	17.733	10.599	30.332	26.676	26.366	23.702
	VDR	1.646	0.930	0.738	0.673	**2.311**	1.926	1.803	1.776	1.504	1.585	1.453	1.426
	PrivBayes	16.247	13.324	11.420	10.736	13.433	12.687	12.058	10.839	21.146	19.640	17.893	17.599
	PrivSTD	**1.364**	**0.415**	**0.298**	**0.127**	3.397	**1.216**	**0.491**	**0.303**	**2.261**	**0.652**	**0.315**	**0.304**
Forecasting Query (sMAPE)	UG	1.551	1.421	1.097	1.073	1.091	0.948	0.829	0.651	1.586	1.351	1.171	1.002
	AG	1.217	1.169	1.150	1.132	0.977	0.789	0.795	0.703	1.228	1.148	1.126	1.092
	HB-Striped	1.731	1.584	1.374	1.284	1.018	0.686	0.533	0.484	1.691	1.536	1.328	1.196
	AHP	1.297	1.259	1.171	1.119	1.607	1.599	1.433	1.196	1.287	1.072	1.069	1.051
	MWEM	1.295	1.236	1.189	1.088	1.379	1.302	1.133	1.027	1.245	1.140	1.004	0.976
	VDR	1.801	1.770	1.747	1.737	1.795	1.789	1.783	1.782	1.930	1.924	1.920	1.915
	PrivBayes	1.293	1.218	1.153	1.069	0.873	0.849	0.854	0.859	1.214	1.053	1.015	1.006
	PrivSTD	**1.175**	**1.159**	**1.068**	**1.025**	**0.657**	**0.458**	**0.435**	**0.421**	**1.199**	**1.025**	**0.943**	**0.925**

Results in terms of Forecasting Query. PrivSTD shows a clear advantage on temporal forecasting under event-level privacy. Most baselines either process time slices independently or capture only limited temporal dependencies, which makes them less effective for long-horizon forecasting. VDR also performs poorly on this task because its architecture is primarily spatial and does not explicitly exploit long-range temporal structure. In contrast, PrivSTD aggregates spectral evidence across time and preserves informative low-frequency components that correspond to persistent temporal patterns. This makes its reconstructions more consistent over long time spans and leads to lower sMAPE across datasets. Under user-level privacy, Forecasting Query is affected by stronger perturbation and the loss of fine-grained trajectory details, but its useful signal differs from hotspot detection. Because forecasting emphasizes smooth temporal evolution and longer-range periodic structure, methods without explicit cross-time structure preservation degrade rapidly as noise accumulates. VDR remains limited by spatial-domain perturbation and weak temporal modeling, whereas PrivSTD benefits from cross-time spectral aggregation.

7.3 Convergence and Runtime

We next analyze convergence and runtime on representative datasets. Figure 5 compares the training convergence of PrivSTD and VDR. Across datasets, PrivSTD converges faster and reaches a lower stabilized relative error, indicating more effective suppression of DP noise, whereas VDR converges more slowly and exhibits larger fluctuations. The higher training error of PrivSTD at the early stage stems from heteroscedastic reconstruction and sparsity-sensitive errors, but these diminish as training stabilizes.

We further examine convergence under the more stringent user-level privacy setting, using Bri_TKY and CABS_SF as examples; similar patterns hold on the remaining datasets. Under user-level privacy, VDR's relative error on Bri_TKY increases with training, indicating severe overfitting to noise. This is because VDR directly adds Laplace noise in the spatial domain, destroying local spatial patterns and hindering its VQ-VAE from learning meaningful

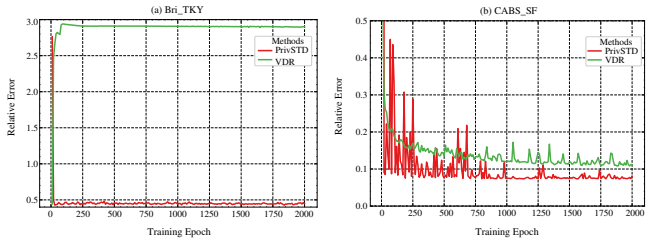

Figure 5: Relative error over training epochs.

spatiotemporal features. Moreover, per-user random sampling introduces two additional challenges: reduced cell counts amplify relative noise in sparse regions, and disrupted trajectory continuity fragments coherent mobility patterns into isolated points. In contrast, PrivSTD captures spatiotemporal features more effectively under high noise via its frequency-domain mechanism, which preserves the underlying density structure, and further leverages CV-R2R to mitigate noise influence. Consequently, it achieves robust and fast convergence in both privacy settings. Table 6 summarizes the end-to-end runtime on four datasets, where Infer denotes the cost of publishing a sanitized dataset. Traditional non-learning mechanisms incur no training cost but require an expensive publishing procedure for each release. Learning-based methods introduce a one-time offline training overhead, after which each release is generated by a single forward pass. Across all datasets, PrivSTD trains faster than VDR, while their inference costs are comparable.

Table 6: Training and inference time (s) across datasets

Method	Dataset							
	4SQ_TKY		CABS_SF		Bri_TKY		Austin	
	Train	Infer	Train	Infer	Train	Infer	Train	Infer
UG	-	0.252	-	0.286	-	0.192	-	0.144
AG	-	3.502	-	3.089	-	0.664	-	0.197
HB-Striped	-	14.364	-	15.164	-	34.679	-	14.656
AHP	-	5.906	-	5.762	-	31.396	-	12.424
MWEM	-	13.776	-	4.818	-	14.827	-	7.056
PrivBayes	-	126.69	-	131.953	-	7.136	-	3.027
VDR	2419.44	0.033	2411.78	0.036	1311.743	0.037	1209.574	0.042
PrivSTD	809.46	0.043	1112.41	0.047	1032.674	0.043	511.496	0.041

Table 5: Utility performance under user-level privacy with varying k_s ($\epsilon = 3$, $H = 384$; best in bold; smaller is better).

Metric	Method	Bri_TKY				Stockholm				Austin			
		$k_s = 3$	$k_s = 9$	$k_s = 15$	$k_s = 21$	$k_s = 3$	$k_s = 9$	$k_s = 15$	$k_s = 21$	$k_s = 3$	$k_s = 9$	$k_s = 15$	$k_s = 21$
Cell Count (RE)	UG	0.480	0.658	0.866	1.057	0.522	0.736	1.108	1.308	0.963	1.811	2.655	3.201
	AG	0.477	0.476	0.475	0.474	0.407	0.404	0.403	0.401	0.641	0.660	0.676	0.718
	HB-Striped	6.326	7.702	8.690	9.569	2.173	2.954	3.628	4.293	2.878	3.542	4.111	4.494
	AHP	0.499	0.494	0.494	0.491	0.420	0.416	0.419	0.421	0.676	0.673	0.686	0.682
	MWEM	0.471	0.483	0.488	0.488	0.411	0.414	0.417	0.418	0.670	0.663	0.663	0.665
	VDR	0.454	0.449	0.445	0.442	0.404	0.398	0.393	0.396	0.620	0.681	0.677	0.674
	PrivBayes	0.495	0.494	0.493	0.494	0.419	0.418	0.420	0.421	0.682	0.683	0.685	0.687
	PrivSTD	**0.414**	**0.421**	**0.420**	**0.420**	**0.387**	**0.378**	**0.363**	**0.366**	**0.554**	**0.655**	**0.642**	**0.628**
Hotspot Query (MAE)	UG	33.415	34.285	37.192	40.709	23.741	22.962	23.186	23.466	16.926	17.453	17.393	17.456
	AG	6.948	5.811	5.980	6.018	8.232	8.131	7.354	7.356	1.759	3.930	4.928	5.441
	HB-Striped	7.038	7.006	6.930	7.022	8.052	8.097	8.087	8.100	1.376	3.332	5.248	5.217
	AHP	48.870	51.636	59.480	59.374	47.401	47.438	47.734	47.834	26.009	31.583	34.048	33.968
	MWEM	11.637	10.208	7.740	8.431	31.063	26.926	27.459	27.358	13.742	18.884	19.826	21.590
	VDR	6.430	6.492	6.472	6.530	**7.453**	**7.415**	7.409	7.301	1.327	3.308	3.683	**3.579**
	PrivBayes	24.885	28.994	19.127	42.340	20.567	24.843	27.720	33.925	3.747	10.002	11.512	12.475
	PrivSTD	**4.762**	**5.439**	**5.327**	**5.413**	33.820	**7.415**	**5.351**	**7.236**	**1.299**	**2.346**	**3.412**	8.576
Forecasting Query (sMAPE)	UG	1.149	0.935	0.951	1.078	1.036	1.025	1.345	1.341	1.348	1.323	1.302	1.290
	AG	0.660	0.734	0.991	1.062	**0.842**	0.975	0.970	0.948	0.996	1.028	1.148	1.086
	HB-Striped	1.551	1.847	1.961	1.872	1.380	1.661	1.624	1.710	1.451	1.376	1.377	1.575
	AHP	1.209	1.244	1.299	1.376	1.213	1.194	1.204	1.207	1.556	1.561	1.572	1.566
	MWEM	1.299	1.158	1.201	1.191	1.189	1.211	1.199	1.221	1.506	1.503	1.525	1.515
	VDR	1.801	1.770	1.747	1.737	1.341	1.250	1.072	0.785	0.762	**0.815**	0.904	0.992
	PrivBayes	0.905	0.820	0.686	0.765	1.208	1.201	1.200	1.154	1.502	1.532	1.520	1.502
	PrivSTD	**0.608**	**0.647**	**0.636**	**0.503**	0.902	**0.839**	**0.755**	**0.764**	**0.667**	0.892	**0.880**	**0.806**

7.4 Parameter Sensitivity Analysis

The default settings used in Figure 6 are highlighted in bold in Table 3. When varying one parameter, all others are fixed. Figure 6 reports the parameter sensitivity results on 4SQ_TKY as a representative example; the complete analysis is provided in the full version [46].

Impact of parameter ϵ. As the privacy budget ϵ increases, the error of most methods consistently decreases across all three query tasks, reflecting the standard privacy–utility trade-off. This effect is more pronounced under user-level privacy, where higher sensitivity makes the low-ϵ regime more challenging. Methods that directly perturb grids or histograms, such as UG, AG, AHP, and HB-Striped, are more sensitive to ϵ, while PrivBayes and MWEM are further limited by model approximation in sparse high-dimensional settings. In contrast, learning-based methods remain relatively robust under small ϵ. In particular, PrivSTD shows the most stable improvement, indicating that frequency-domain compression and denoising are especially effective for preserving spatio-temporal structure.

Impact of parameter $\mathbb{H}$. Increasing H produces finer but sparser grids: per-cell counts decrease while the relative effect of privacy noise grows. Consequently, Hotspot and Forecasting Query errors generally increase, whereas Cell Count often follows a decrease-then-increase trend as reduced discretization bias gives way to amplified perturbation noise. This effect is strongest for UG and AG and also destabilizes hierarchical and clustering-based methods at high resolution. VDR partly mitigates the degradation, but preserving utility at large H remains difficult without stable spatio-temporal patterns.

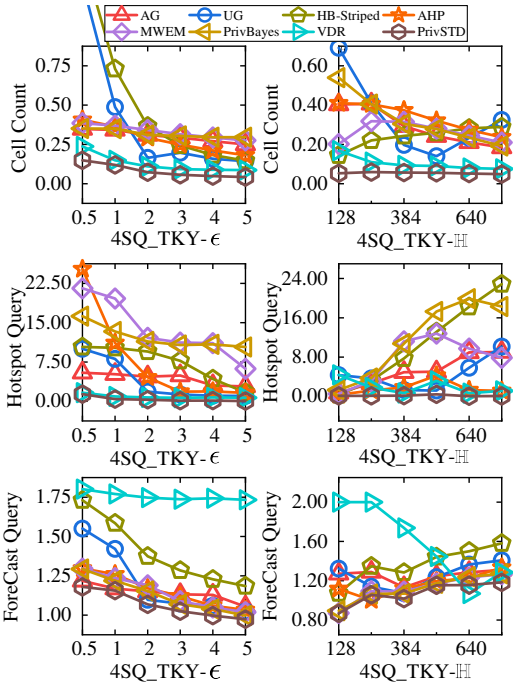

Figure 6: Parameter sensitivity on the 4SQ_TKY dataset.

8 CONCLUSION

We presented PrivSTD, a two-stage method for differentially private high-resolution spatio-temporal density publication. It combines privacy-guaranteed frequency estimation, which preserves dominant spatial structure and suppresses noise-dominated components, with covariance-aware self-supervised denoising, which reduces structured heteroscedastic noise without clean data. Experiments on real-world datasets show that PrivSTD consistently outperforms state-of-the-art baselines.

ACKNOWLEDGMENTS

This work was supported in part by the NSFC under Grant Nos. 62472377, 62402431, and 62441618, and by the Zhejiang Provincial Natural Science Foundation of China under Grant No. Z26F020046. Yangyang Wu is the corresponding author.

REFERENCES

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the ACM SIGSAC conference on computer and communications security*. 308–318.

[2] Gergely Acs and Claude Castelluccia. 2014. A case study: Privacy-preserving release of spatio-temporal density in Paris. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*. Association for Computing Machinery, 1679–1688.

[3] N. Ahmed, T. Natarajan, and K. R. Rao. 1974. Discrete cosine transform. *IEEE Trans. Comput.* 23, 1 (Jan. 1974), 90–93.

[4] Ritesh Ahuja, Sepanta Zeighami, Gabriel Ghinita, and Cyrus Shahabi. 2023. A neural approach to spatio-temporal data release with user-level differential privacy. *Proceedings of the ACM on Management of Data* 1, 1 (2023), 1–25.

[5] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.

[6] Yoav Benjamini and Daniel Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of statistics* (2001), 1165–1188.

[7] George Casella and Roger Berger. 2024. *Statistical inference*. Chapman and Hall/CRC.

[8] Kamalika Chaudhuri, Claire Monteleoni, and Anand D Sarwate. 2011. Differentially private empirical risk minimization. *Journal of Machine Learning Research* 12, 3 (2011).

[9] Rui Chen, Qian Xiao, Yu Zhang, and Jianliang Xu. 2015. Differentially private high-dimensional data publication via sampling-based inference. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*. 129–138.

[10] Shi Chen, Daniel Janies, Rajib Paul, and Jean-Claude Thill. 2024. Leveraging advances in data-driven deep learning methods for hybrid epidemic modeling. *Epidemics* 48 (2024), 100782.

[11] Shiyan Chen, Jiyuan Zhang, Zhaofei Yu, and Tiejun Huang. 2024. Exploring efficient asymmetric blind-spots for self-supervised denoising in real-world scenarios. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2814–2823.

[12] Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the ACM SIGKDD international conference on Knowledge discovery and data mining*. 1082–1090.

[13] Federico Delussu, Michele Tizzoni, and Laetitia Gauvin. 2023. The limits of human mobility traces to predict the spread of COVID-19: A transfer entropy approach. *PNAS nexus* 2, 10 (2023), pgad302.

[14] Bangchao Deng, Ling Ding, Lianhua Ji, Chunhua Chen, Xin Jing, Bingqing Qu, and Dingqi Yang. 2025. Marionette: Fine-grained conditional generative modeling of spatiotemporal human trajectory data beyond imitation. In *Proceedings of the ACM SIGKDD conference on knowledge discovery and data mining*. 463–473.

[15] Milos Doroslovacki and Hong Fan. 1993. Wavelet-based adaptive filtering. In *IEEE international conference on acoustics, speech, and signal processing*, Vol. 3. 488–491.

[16] Cynthia Dwork. 2006. Differential privacy. In *ICALP*. Springer-Verlag, Berlin, Heidelberg, 1–12.

[17] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*. 265–284.

[18] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N Rothblum. 2010. Differential privacy under continual observation. In *Proceedings of the ACM symposium on theory of computing*. 715–724.

[19] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and trends® in theoretical computer science* 9, 3–4 (2014), 211–407.

[20] Facebook Data for Good. 2022. Facebook Data for Good: High Resolution Density Maps. https://dataforgood.facebook.com/dfg/tools/highresolution-population-density-maps. Accessed: October 10, 2025.

[21] Moritz Hardt, Katrina Ligett, and Frank McSherry. 2012. A simple and practical algorithm for differentially private data release. *Advances in neural information processing systems* 25 (2012).

[22] Danlei Hu, Lu Chen, Hanxi Fang, Ziquan Fang, Tianyi Li, and Yunjun Gao. 2023. Spatio-temporal trajectory similarity measures: A comprehensive survey and quantitative study. *IEEE Transactions on Knowledge and Data Engineering* 36, 5 (2023), 2191–2212.

[23] Daniel Im Im, Sungjin Ahn, Roland Memisevic, and Yoshua Bengio. 2017. Denoising criterion for variational auto-encoding framework. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 31.

[24] Shankar Iyer, Brian Karrer, Daniel T Citron, Farshad Kooti, Paige Maas, Zeyu Wang, Eugenia Giraudy, Ahmed Medhat, P Alex Dow, and Alex Pompe. 2023. Large-scale measurement of aggregate human colocation patterns for epidemiological modeling. *Epidemics* 42 (2023), 100663.

[25] Daniel Kifer, Adam Smith, and Abhradeep Thakurta. 2012. Private convex empirical risk minimization and high-dimensional regression. In *conference on learning theory*. 25–1.

[26] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. 2019. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2129–2137.

[27] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. 2018. Noise2Noise: Learning image restoration without clean data. In *international conference on machine learning*. 4620–4631.

[28] Ryan Mckenna, Daniel Sheldon, and Gerome Miklau. 2019. Graphical-model based estimation and inference for differential privacy. In *PMLR*, Vol. 97. 4435–4444.

[29] Hermine Lore Nguena Nguefack, M Gabrielle Pagé, Joel Katz, Manon Choinière, Alain Vanasse, Marc Dorais, Oumar Mallé Samb, and Anaïs Lacasse. 2020. Trajectory modelling techniques useful to epidemiological research: a comparative narrative review of approaches. *Clinical epidemiology* (2020), 1205–1222.

[30] Behnam Nikparvar, Md Mokhlesur Rahman, Faizeh Hatami, and Jean-Claude Thill. 2021. Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network. *Scientific reports* 11, 1 (2021), 21715.

[31] Alan V Oppenheim. 1999. *Discrete-time signal processing*. Pearson Education India.

[32] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. 2021. R2R: Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2043–2052.

[33] Michal Piorkowski, Natasa Sarafijanovic-Djukic, and Matthias Grossglauser. 2009. CRAWDAD data set epfl/mobility (v. 2009-02-24).

[34] DD-Y Po and Minh N Do. 2006. Directional multiscale modeling of images using the contourlet transform. *IEEE Transactions on image processing* 15, 6 (2006), 1610–1620.

[35] Wahbeh Qardaji, Weining Yang, and Ninghui Li. 2013. Differentially private grids for geospatial data. In *2013 IEEE 29th international conference on data engineering*. IEEE, 757–768.

[36] Wahbeh Qardaji, Weining Yang, and Ninghui Li. 2013. Understanding hierarchical methods for differentially private histograms. *Proceedings of the VLDB Endowment* 6, 14 (2013), 1954–1965.

[37] K Ramamohan Rao and Ping Yip. 2014. *Discrete cosine transform: algorithms, advantages, applications*. Academic press.

[38] Vibhor Rastogi and Suman Nath. 2010. Differentially private aggregation of distributed time-series with transformation and encryption. In *Proceedings of the ACM SIGMOD international conference on management of data*. 735–746.

[39] John D Storey. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 64, 3 (2002), 479–498.

[40] Hien To, Gabriel Ghinita, Liyue Fan, and Cyrus Shahabi. 2016. Differentially private location protection for worker datasets in spatial crowdsourcing. *IEEE Transactions on Mobile Computing* 16, 4 (2016), 934–949.

[41] Jonathan Ullman and Adam Sealfon. 2019. Efficiently estimating erdos-renyi graphs with node differential privacy. *Advances in Neural Information Processing Systems* 32 (2019).

[42] Behzad Vahedi, Morteza Karimzadeh, and Hamidreza Zoraghein. 2021. Spatiotemporal prediction of COVID-19 cases using inter-and intra-county proxies of human interactions. *Nature communications* 12, 1 (2021), 6440.

[43] Xi Wu, Fengan Li, Arun Kumar, Kamalika Chaudhuri, Somesh Jha, and Jeffrey Naughton. 2017. Bolt-on differential privacy for scalable stochastic gradient descent-based analytics. In *Proceedings of the 2017 ACM international conference on management of data*. 1307–1322.

[44] Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. 2010. Differential privacy via wavelet transforms. *IEEE Transactions on knowledge and data engineering* 23, 8 (2010), 1200–1214.

[45] Dingqi Yang, Bingqing Qu, Jie Yang, and Philippe Cudre-Mauroux. 2019. Revisiting user mobility and social relationships in LBSNs: A hypergraph embedding approach. In *The world wide web conference*. 2147–2157.

[46] Shuzhan Ye, Yujia Hu, Lu Chen, Yangyang Wu, Zhikun Zhang, Tianyi Li, and Christian S. Jensen. 2026. PrivSTD: Differentially Private Spatio-temporal Trajectory Density Data Publication. (2026). https://github.com/YeShuzhan/PrivSTD/blob/main/PrivSTD.pdf Accessed: June 19, 2026.

[47] Sepanta Zeighami, Ritesh Ahuja, Gabriel Ghinita, and Cyrus Shahabi. 2022. A neural database for differentially private spatial range queries. *Proceedings of the VLDB Endowment* 15, 5 (2022), 1066–1078.

[48] Ang Zhang, Aqil Tariq, Abdul Quddoos, Iram Naz, Rana Waqar Aslam, Elgar Barboza, Sajid Ullah, and M Abdullah-Al-Wadud. 2025. Spatio-temporal analysis of urban expansion and land use dynamics using google earth engine and predictive models. *Scientific Reports* 15, 1 (2025), 6993.

[49] Jun Zhang, Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, and Xiaokui Xiao. 2017. PrivBayes: Private data release via Bayesian networks. *ACM Trans. Database Syst.* 42, 4, Article 25 (2017), 41 pages.

[50] Xiaojian Zhang, Rui Chen, Jianliang Xu, Xiaofeng Meng, and Yingtao Xie. 2014. Towards accurate histogram publication under differential privacy. In *Proceedings of the SIAM international conference on data mining*. SIAM, 587–595.

[51] Zhikun Zhang, Tianhao Wang, Ninghui Li, Jean Honorio, Michael Backes, Shibo He, Jiming Chen, and Yang Zhang. 2021. PrivSyn: Differentially private data synthesis. In *USENIX Security*.
[52] Dihan Zheng, Sia Huat Tan, Xiaowen Zhang, Zuoqiang Shi, Kaisheng Ma, and Chenglong Bao. 2020. An unsupervised deep learning approach for real-world image denoising. In *international conference on learning representations*.
[53] Andreas Züfle, Dieter Pfoser, Carola Wenk, Andrew Crooks, Hamdi Kavak, Taylor Anderson, Joon-Seok Kim, Nathan Holt, and Andrew Diantonio. 2024. In Silico Human Mobility Data Science: Leveraging Massive Simulated Mobility Data. *ACM Transactions on Spatial Algorithms and Systems* 10, 2 (2024), 1–27.