MCM-DM: Towards Better Spatio-Temporal Event Representation Learning via Discrete Morse Theory

Yuanheng Zhang*
Department of Statistics
University of California, Riverside
yzhan1242@ucr.edu

Jennifer Rozenblit*
Department of Mathematics
University of Texas at Austin
jrozenblit@utexas.edu

Chenguang Yang
Department of Statistics
University of California, Riverside
cyang314@ucr.edu

Jaidev Goel
Department of Statistics
Virginia Tech
jaidev@vt.edu

Yulia R. Gel
Department of Statistics
Virginia Tech
ygl@vt.edu

Yuzhou Chen
Department of Statistics
University of California, Riverside
yuzhou.chen@ucr.edu

Abstract

Spatio-temporal point processes (STPPs) represent a random collection of points where each point corresponds to the time and location of an event. STPPs are widely used to describe a wide range of phenomena, from earthquakes to wildfires to crime occurrence to disease outbreaks. Generative models have recently emerged as a new powerful paradigm for STPPs, demonstrating highly competitive generalization capabilities and promising potential for systematic uncertainty quantification. However, existing generative approaches primarily rely on unimodal numerical data, overlooking the geographic context in which events occur and failing to capture critical spatial and temporal patterns within the STPP. To address these limitations, we propose Morse-aware Cross-Modality learning within Diffusion Model, or **MCM-DM**. MCM-DM represents a cross-modality diffusion framework based on the discrete Morse and cobordism theories that augments STPP modeling with geographic scene understanding. MCM-DM consists of three key components: (i) a vision-language model-based geographic encoder that extracts semantic embeddings from satellite tiles at each event location; (ii) an attention-based fusion mechanism to integrate critical structure representation with spatio-temporal embeddings, and (iii) a Morse-theoretic topological aligner which aligns latent representations of critical events with spatio-temporal embedding space. We demonstrate the MCM-DM utility on 8 diverse STPP datasets from Earth sciences, epidemiology, urban mobility, and crime analytics, highlighting its cross-domain versatility. The code is available at https://github.com/Evanzhang100302/MCM-DM.

1 Introduction

Spatio-temporal point processes (STPPs) provide a stochastic framework for modeling asynchronous event data, in which both the event times and locations are treated as random variables of primary

* Joint first authors.

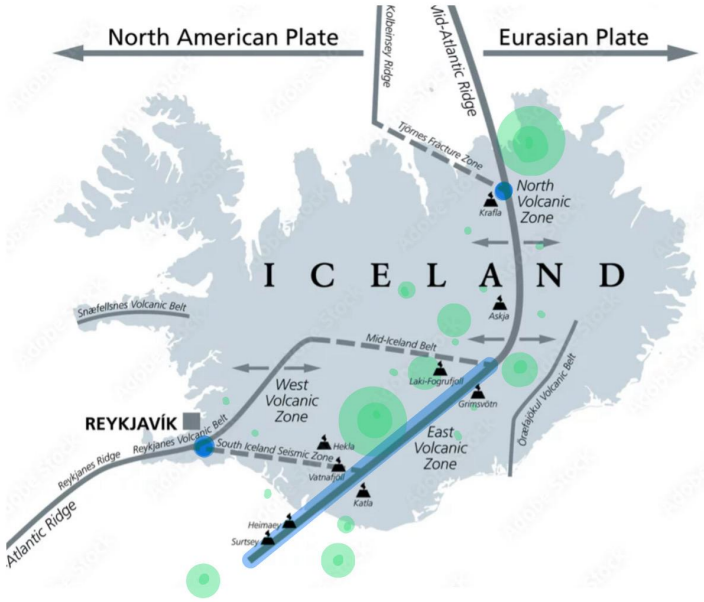
40th Conference on Neural Information Processing Systems (NeurIPS 2026).

interest. STPPs are widely adopted as the primary abstraction for describing a variety of phenomena, including earthquake aftershocks [36, 37], the spread of infectious disease such as COVID-19 [7], and urban crime [33]. Further, STPPs have long-standing history in statistical science, with classical approaches, such as inhomogeneous Poisson, Hawkes, Neyman–Scott, and Cox models, specifying the conditional intensity in closed form [8, 9, 35]. However, assumptions of parametric kernels limit model expressiveness, and a substantial line of recent neural and generative STPPs has emerged to lift these restrictions, ranging from neural Hawkes and Transformer-Hawkes models [12, 30, 60] to diffusion-based generative formulations such as DSTPP [54]. Despite this progress, most prevailing neural and generative STPPs operate on numerical timestamp-pair event tuples alone, and discard the essential information on the geographic context in which those coordinates are embedded. This constitutes a substantive omission. For example, earthquake aftershocks are shaped by fault geometry; urban crime and traffic incidents are shaped by infrastructure and the built environment. These details are not recoverable from a latitude-longitude pair, yet all of it is visible in a satellite tile centered at that pair. A model that ignores the geographic scene is, therefore, working with a suppressed view of the data-generating process.

Recent vision-language models (VLMs) trained on remote-sensing data [26, 25, 47, 51, 52] now extract land-use, infrastructure, and terrain semantics directly from satellite imagery, explicitly providing the geographic prior that numerical STPPs lack. These capabilities suggest an obvious recipe: encode each event's local satellite tile with a pretrained VLM, fuse the resulting embedding with the spatial and temporal embeddings, and condition a generative STPP on the joint representation. The difficulty lies in the alignment, since geographic embeddings produced by a VLM live in a representation space optimized for a generic image-text contrastive objective, while spatio-temporal embeddings produced by an STPP backbone live in a space organized by event dynamics. Concatenating or cross-attending between the two modalities couples them statistically but provides no *structural anchor* that tells the model which events in the sequence are the ones whose geographic context most matters. Without such an anchor, the geographic modality is at best a regularizer and at worst a confound; that is, the model has no principled way to attend to the satellite tile of an aftershock at a fault intersection more than to the tile of a routine background event.

A natural combinatorial object associated with an STPP realization is a graph on the events themselves. Indeed, events are discrete, indexed, and related through local space-time conditioning — the conditional intensity at one event depends on the recent history of nearby events — which is precisely the structure a graph encodes. Furthermore, graph abstraction allows for more flexible and systematic treatment of various higher-order interdependencies within STPP that typically cannot be rigorously captured otherwise [22, 27]. Several constructions can serve this purpose, including ε-neighborhood graphs, Delaunay triangulations, Gabriel graphs, and k-nearest-neighbors (kNN) graphs, each making different trade-offs between locality, sparsity, and adaptivity to the underlying density. We focus on the kNN graph abstraction under a normalized space-time metric, which adapts to the heterogeneous event densities characteristic of STPPs and yields a bounded-degree combinatorial summary of the realized correlation structure of the sequence without commitment to any parametric form for the intensity. Given this graph representation, the next question is which events are likelier to constitute more to the structural backbone of the graph. We answer this through *discrete Morse theory* [13, 14], which identifies the cells of a complex at which the topology changes - the *critical cells*.

A pair of discrete Morse functions, one spatial and one causally temporal, partitions the edges of the event graph into four populations according to whether each edge is critical or paired under each function: spatio-temporal recurrences, persistent hotspots, traveling waves, and a tree-like backbone. The events incident to the non-backbone populations are the topologically distinguished ones, and represent the anchors the cross-modal aligner attends to. Intuitively, as shown in Fig. 1, tectonic plates constitute a Morse-backbone for earthquake events (green). The critical points in these plate interactions constitute critical cells (blue). **Spatial and Temporal Morse Functions.** Armed with these ideas, we propose **MCM-DM** (Morse-aware Cross-Modality learning with Diffusion Model), a diffusion-based generative STPP that conditions on a fused representation from (i) spatio-temporal learner, (ii) Morse-based spatio-temporal learner, (iii) VLM-derived geographic embeddings of event-centered satellite imagery. A Morse-guided aligner uses the Morse decomposition to register critical events with the spatio-temporal embedding space. We prove that the skeleton component of this construction, namely the four-way Morse decomposition of the event graph, is exactly invariant under raw data perturbations smaller than a computable per-realization stability radius. We derive the



stability bound and show empirically that MCM-DM outperforms state-of-the-art generative STPP baselines on 8 real-world datasets. Significance of our contributions can be summarized as follows:

- We introduce a discrete-Morse-theoretic decomposition of the STPP event graph abstraction, using a pair of spatial and temporal Morse functions whose intersection yields a four-way partition of edges into spatio-temporal recurrences, persistent hotspots; we identify the geographic-modality gap in neural and generative STPPs and propose a cross-modal diffusion framework that closes it by integrating VLM-derived satellite-tile embeddings with spatio-temporal representations.
- We develop a Morse-theoretic aligner that registers critical events with the spatio-temporal embedding space. We prove that the four-way Morse decomposition of the event graph is exactly invariant under raw-data perturbations smaller than a computable per-realization radius, certifying stability of the topological component of the conditioning signal.
- We show that MCM-DM is versatile, effective and efficient, consistently outperforming state-of-the-art baselines by up to **196.55%** and **91.11%** in spatial and temporal dimensions respectively. Additionally, it achieves up to **3.53x** faster inference compared to runner-ups.

2 Related Works

Diffusion for Spatio-Temporal Point Processes (STPPs). STPPs characterize asynchronous events through a conditional intensity function over time and space, with classical formulations rooted in (in)homogeneous Poisson processes [9], the Hawkes process [18], the Neyman-Scott process [35], and Cox processes [8, 10]. STPPs have been widely studied in machine learning, including kernel-based methods, normalizing flows, and neural ODEs, just to name a few (see [6, 34, 42] for comprehensive overviews and Appendix H for more detail). Furthermore, generative formulations of STPPs have rapidly advanced, particularly, focusing on diffusion approaches [28, 40, 53, 54, 57]. Despite their premise, especially in terms of uncertainty quantification most diffusion methods for STPPs rely solely on numerical timestamps and coordinates and ignore the geographic semantics in which events occur.

Discrete Morse Theory in Machine Learning. Classical Morse theory provides a fundamentally geometric description of how gradient flow organizes a manifold into cells via stable and unstable manifolds [32, 43]. Forman's discrete Morse theory [13, 14] extends these ideas to cell complexes. Recent results of [24, 50] show the utility of discrete Morse theory in application to reductions of persistent homology computations on graphs, while [29] studies it in the context of topological quantum field theory. Besides this, discrete Morse theory still remains largely confined to pure mathematics and its potential in machine learning is yet untapped.

3 Methodology: MCM-DM

Problem Setup. We are given a sequence of events $\boldsymbol{\mathcal{X}} = \{\boldsymbol{x}_i \mid i = 1, 2, \ldots, N\}$ over an observation window $[0, T]$, where N is the number of events, and each event $\boldsymbol{x}_i = (t_i, s_i)$ contains an event timestamp $t_i \in \mathbb{R}$ and a geographical coordinate $s_i \in \mathbb{R}^d$ (typically $d = 2$, i.e., latitude and longitude). To effectively capture dependencies among events, we use a normalized joint spatio-temporal distance as $d(x_i, x_j) = \sqrt{\frac{(t_i - t_j)^2}{\sigma_t^2} + \frac{\|s_i - s_j\|^2}{\sigma_s^2}}$, where $\sigma_t, \sigma_s > 0$ are bandwidth parameters selected by cross-validation or Silverman's rule on the temporal and spatial marginals respectively. The *event graph* $G = (V, E)$ is the undirected kNN graph under d with $k = \lceil \sqrt{n} \rceil$:

$$V = \mathcal{S}, \quad E = \big\{\{x_i, x_j\} : x_j \in \mathrm{kNN}_k(x_i) \text{ or } x_i \in \mathrm{kNN}_k(x_j)\big\}, \tag{1}$$

where $\mathcal{S} = \{s_1, \ldots, s_n\}$. The symmetrization ensures G is undirected, while the choice $k = \lceil \sqrt{n} \rceil$ balances local connectivity and preserves essential topological structures (e.g., β_1) against sparsity, which keeps computation tractable. We first recall the necessary background, then define the two Morse functions and their partition.

3.1 Discrete Morse Functions on Graphs

A *discrete Morse function* on a graph $G = (V, E)$ assigns a real value $g(v)$ to every node and $g(e)$ to every edge such that the resulting *gradient field*, the set of node-edge pairs (v, e) with v a face of e and $g(v) \geq g(e)$, is acyclic [13] (for further background, refer to Appendix B). Any such gradient field partitions the edge set into two disjoint populations as $E = E^{\mathrm{crit}} \sqcup E^{\mathrm{pair}}$, where a *paired edge* $e \in E^{\mathrm{pair}}$ is matched with exactly one of its endpoint nodes (forming an aforementioned gradient pair), and a *critical edge* $e \in E^{\mathrm{crit}}$ is an edge that remains unpaired. Due to the acyclicity inherent from the gradient axioms in discrete Morse theory, paired edges always form a forest, and therefore each critical edge closes exactly one fundamental cycle in G relative to this forest. In practice, we construct the gradient field via a value-monotone edge-iteration procedure following [24], restricted to graphs. A single Morse function on the joint space-time event graph treats spatial and temporal structure symmetrically, with both entering through a shared kernel density estimate over a normalized (t, s) metric. While density maxima in $\mathbb{R}^2$ naturally create cycles in a kNN graph, this construction would favor spatial over temporal structure for reasons rooted in causality, dimension, and dynamics. STPPs are causal processes, in the sense that only past events influence the present ones via a conditional intensity. Yet, a symmetric kernel treats past and future neighbors identically, conflating causal influence with temporal coincidence. Since time is one-dimensional, the k temporally nearest events form an approximate path graph, contributing chain structure (absorbed into E^{pair}) rather than cycles (which would appear in E^{crit}), whereas two-dimensional space creates loops more readily and dominates the cycle structure. Finally, the temporal phenomena of interest, such as self-excitation, burstiness, and periodicity, are properties of inter-event intervals and ordering rather than of static density, so while the spatial kernel captures clusters (which are density maxima), the temporal kernel misses directional, dynamical patterns. Applying two separate Morse functions to the same graph, one spatial and one temporal, preserves the joint space-time coupling encoded in the graph topology while mitigating these asymmetries. The spatial Morse function is defined as:

$$g_s(x_i) = -\sum_{j \in \mathcal{N}(x_i)} \exp\!\left(-\frac{\|s_i - s_j\|^2}{2\sigma_s^2}\right) + \varepsilon_i^{(s)}, \tag{2}$$

where $\varepsilon_i^{(s)} \sim \mathrm{Unif}(0, \delta)$ is a small tie-breaking jitter and $\mathcal{N}(x_i)$ is direct neighbors of x_i in G. The sum is symmetric over all graph neighbors regardless of temporal ordering. Note that high density areas point to large kernel sums, and hence result in large negative scores or low Morse values, while spatially isolated events receive high values and become critical nodes. Further, we define the temporal Morse function:

$$g_t(x_i) = -\sum_{\substack{j \in \mathcal{N}(x_i) \\ t_j < t_i}} \exp\!\left(-\frac{(t_i - t_j)^2}{2\sigma_t^2}\right) + \varepsilon_i^{(t)}. \tag{3}$$

The sum runs only over neighbors x_j with $t_j < t_i$, which is the *causal past* of event x_i. The key asymmetry lies in that $g_t(x_i)$ depends only on the history $\mathcal{H}_{t_i} = \{x_j : t_j < t_i\}$, mirroring the

conditional intensity. Events that follow a dense burst of recent past activity receive low Morse values, while events arriving after a long quiet period such as regime changes, surprise onsets, receive high values and become temporally critical. Each Morse function induces edge values by 'extending' its node values via $g_s(e) = \frac{1}{2}(g_s(x_i) + g_s(x_j))$, $g_t(e) = \frac{1}{2}(g_t(x_i) + g_t(x_j))$ and greedy Forman pairing (Section 3.1) is applied independently to each function to produce two decompositions of the same edge set (further expanded upon in Appendix D):

$$E = E_s^{\text{crit}} \sqcup E_s^{\text{pair}} \qquad \text{and} \qquad E = E_t^{\text{crit}} \sqcup E_t^{\text{pair}}. \tag{4}$$

The intersection of these two decompositions yields the four-way partition

$$E = E_{cc} \sqcup E_{cp} \sqcup E_{pc} \sqcup E_{pp},$$

where $E_{cc} = E_s^{\text{crit}} \cap E_t^{\text{crit}}$, $E_{cp} = E_s^{\text{crit}} \cap E_t^{\text{pair}}$, $E_{pc} = E_s^{\text{pair}} \cap E_t^{\text{crit}}$, $E_{pp} = E_s^{\text{pair}} \cap E_t^{\text{pair}}$. We call these four populations spatio-temporal recurrences, persistent spatial hotspots, traveling-wave edges, and backbone edges, respectively.

3.2 Choice of Morse Function

Rather than enumerating discrete Morse functions, we adopt the *Matérn family*, a one-parameter family of kernels parameterized by smoothness $\nu > 0$ that interpolates between the heavy-tailed exponential and the smooth Gaussian. This family is the canonical choice in spatial statistics and Gaussian-process modeling [38, 45] and yields decision-gap Lipschitz constants that vary smoothly with ν. We then describe two non-separable variants that introduce explicit space-time coupling.

Definition 1 (Matérn kernel)**.** For smoothness $\nu > 0$, the (unit-bandwidth) Matérn kernel is

$$K_\nu(r) \;=\; \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\, r\right)^{\nu} \mathrm{K}_\nu\!\left(\sqrt{2\nu}\, r\right), \tag{5}$$

where Γ is the Gamma function and K_ν is the modified Bessel function of the second kind of order ν.

The smoothness parameter ν controls regularity precisely: K_ν induces $\lceil \nu \rceil - 1$ times mean-square differentiable random fields [45]. For different values of ν, K_ν recovers several standard kernels widely used in spatial statistics and geostatistics; Table 4 lists several closed forms at the half-integer.

Definition 2 (Matérn Morse Functions)**.** Replacing the Gaussian in (2)–(3) with K_ν yields the *Matérn Morse functions*

$$g_s^{(\nu)}(x_i) \;=\; -\sum_{j \in \mathcal{N}(x_i)} K_\nu\!\left(\tfrac{\|s_i - s_j\|}{\sigma_s}\right) + \varepsilon_i^{(s)} \qquad g_t^{(\nu)}(x_i) \;=\; -\sum_{\substack{j \in \mathcal{N}(x_i) \\ t_j < t_i}} K_\nu\!\left(\tfrac{|t_i - t_j|}{\sigma_t}\right) + \varepsilon_i^{(t)}. \tag{6}$$

3.3 Non-Separable Space-Time Coupling

Separable kernels treat space and time as independent axes, but many real STPPs exhibit coupling, as in diffusive epidemic spread where the spatial scale of new cases grows with elapsed time. We consider two non-separable alternatives that retain the Matérn smoothness parameterization.

Matérn–Mahalanobis kernel. Replacing the diagonal metric of $d(x_i, x_j)$ with a full $(d+1) \times (d+1)$ covariance Σ and applying K_ν to the Mahalanobis distance gives

$$K_\nu^{\mathrm{M}}(x_i, x_j) \;=\; K_\nu\big(d_M(x_i, x_j)\big), \quad d_M(x_i, x_j) \;:=\; \sqrt{(x_i - x_j)^\top \Sigma^{-1} (x_i - x_j)}. \tag{7}$$

Off-diagonal entries encode preferred space-time alignment, with large variance along a $(\delta t, \delta s)$ direction corresponding to a known spreading velocity. The induced Morse function is symmetric in time, yielding a single decomposition $E = E_{\mathrm{M}}^{\text{crit}} \sqcup E_{\mathrm{M}}^{\text{pair}}$ rather than the four-way partition, but in exchange encodes directional structure.

Matérn–Hawkes coupled kernel. A causal kernel whose spatial bandwidth grows with elapsed time:

$$K_\nu^{\mathrm{hw}}(x_i, x_j) \;=\; \mathbf{1}\{t_j < t_i\}\, K_\nu\!\left(\tfrac{|t_i - t_j|}{\tau}\right) K_\nu\!\left(\tfrac{\|s_i - s_j\|}{\sigma(t_i - t_j)}\right), \tag{8}$$

with $\sigma^2(\delta t) = \sigma_0^2 + \kappa\, \delta t$ for diffusion coefficient $\kappa \geq 0$. When $\kappa = 0$ it factorises into a separable causal Matérn; when $\kappa > 0$ it captures diffusive spread. The raw-data skeleton stability theorem applies to any ν, with L_s, L_t replaced by the corresponding ν-dependent constants. The stability radius is a smooth function of ν, so smoothness can be tuned alongside the bandwidths (σ_t, σ_s) in the same stability-aware hyperparameter sweep.

4 Theoretical Framework: Skeleton, Perturbations, Stability

This section establishes the formal apparatus needed to state stability results for the four-way Morse decomposition. We name the four-way partition as a single object (the *event skeleton*), define a perturbation model adapted to STPP data, and introduce a decision margin that yields a computable stability certificate.

Definition 3 (Event skeleton)**.** Let $G = (V, E)$ be the joint kNN event graph of an STPP realization $\mathcal{S} = \{x_1, \ldots, x_L\}$, and let g_s, g_t be the spatial and causal-temporal Morse functions of (2) and (3). The *event skeleton* of $\mathcal{S}$ is the four-way edge partition $\mathrm{Skel}(\mathcal{S}) = \big(E_{cc}, E_{cp}, E_{pc}, E_{pp}\big)$ together with the assignment of each edge to its quadrant, where:

$$E_{cc} = E_s^{\mathrm{crit}} \cap E_t^{\mathrm{crit}},\ E_{cp} = E_s^{\mathrm{crit}} \cap E_t^{\mathrm{pair}},\ E_{pc} = E_s^{\mathrm{pair}} \cap E_t^{\mathrm{crit}},\ E_{pp} = E_s^{\mathrm{pair}} \cap E_t^{\mathrm{pair}}.$$

We call E_{cc} spatio-temporal recurrences, E_{cp} persistent spatial hotspots, E_{pc} traveling-wave edges, and E_{pp} backbone edges.

The skeleton $\mathrm{Skel}(\mathcal{S})$ is the canonical topological summary of a realization, and is the conditioning signal consumed by MCM-DM's denoising network. All stability statements below concern invariance or controlled change of $\mathrm{Skel}(\mathcal{S})$ under perturbations of $\mathcal{S}$, (σ_t, σ_s), or the ambient kernel. The event skeleton encodes *flow-induced* edge classification, tracking which cycle each critical edge closes and which spanning-forest path each paired edge belongs to. We work with three nested classes of perturbations, each with its own stability question.

Definition 4 (STPP-data perturbation)**.** A *data perturbation* of $\mathcal{S}$ is a map $\mathcal{S} \mapsto \mathcal{S}'$ where $\mathcal{S}' = \{x'_1, \ldots, x'_L\}$ with $x'_i = (t'_i, s'_i)$ and, for perturbation magnitude is $\rho \geq 0$,

$$\max_i \sqrt{\frac{(t_i - t'_i)^2}{\sigma_t^2} + \frac{\|s_i - s'_i\|^2}{\sigma_s^2}} \ \leq\ \rho.$$

Definition 5 (Bandwidth perturbation)**.** A *bandwidth perturbation* replaces (σ_t, σ_s) by (σ'_t, σ'_s) with $\max\{|\sigma_t - \sigma'_t|/\sigma_t, |\sigma_s - \sigma'_s|/\sigma_s\} \leq \rho$. The event set $\mathcal{S}$ and graph G are unchanged but the Morse values $g_s(\cdot), g_t(\cdot)$ on each node shift. We say that a perturbation is *skeleton-preserving* if $\mathrm{Skel}(\mathcal{S}) = \mathrm{Skel}(\mathcal{S}')$.

Definition 6 (Decision margin and structural margins)**.** For each adjacent node pair $\{x_i, x_j\}$ in G, define the spatial and temporal decision gaps $\delta_{ij}^{(s)}(\mathcal{S}) := g_s(x_i) - g_s(x_j)$, and $\delta_{ij}^{(t)}(\mathcal{S}) := g_t(x_i) - g_t(x_j)$, and let $J = \{(i,j) : \{x_i, x_j\} \in E\}$. Assuming the no-ties condition $\delta_{ij}^{(s)}, \delta_{ij}^{(t)} \neq 0$ for all $(i,j) \in J$, the *skeleton decision margin* is $m(\mathcal{S}) := \min_{(i,j)\in J} \min\Big\{\big|\delta_{ij}^{(s)}(\mathcal{S})\big|,\ \big|\delta_{ij}^{(t)}(\mathcal{S})\big|\Big\}$.

Let $d_i^{(1)}(\mathcal{S}) \leq \cdots \leq d_i^{(L-1)}(\mathcal{S})$ denote the ordered bandwidth-normalized distances from x_i to all other events. The *k-nearest-neighbor margin* and *temporal-order margin* are

$$\eta_{k\mathrm{NN}}(\mathcal{S}) := \min_i \big(d_i^{(k+1)}(\mathcal{S}) - d_i^{(k)}(\mathcal{S})\big), \quad \eta_{\mathrm{ord}}(\mathcal{S}) := \min_{i \neq j} \frac{|t_i - t_j|}{\sigma_t}.$$

The no-ties condition is generic and is achievable by the infinitesimal jitter $\varepsilon_i^{(s)}, \varepsilon_i^{(t)}$ already included in the Morse-function definitions. The structural margins $\eta_{k\mathrm{NN}}$ and η_{ord} are positive on generic STPP realizations but can vanish in datasets with discretized timestamps or coincident events; we discuss this edge case in Appendix E.5.

Definition 7 (Raw-data stability radius)**.** The *raw-data stability radius* of $\mathcal{S}$ is $r_{\mathrm{raw}}(\mathcal{S}) := \min\{\eta_{k\mathrm{NN}}(\mathcal{S})/4,\ \eta_{\mathrm{ord}}(\mathcal{S})/2,\ m(\mathcal{S})/L(\nu)\}$, where $L(\nu) = 4\|K'_\nu\|_\infty A_{\max}$ is the unified decision-gap Lipschitz constant from Eq. (15).

4.1 Stability Theorem

Theorem 1 (Raw-data perturbation stability of the event skeleton)**.** *Suppose $\mathcal{S}$ satisfies the no-ties condition of Definition 23 and the structural-margin conditions $\eta_{k\mathrm{NN}}(\mathcal{S}) > 0$ and $\eta_{\mathrm{ord}}(\mathcal{S}) > 0$. Suppose further that the spatial and temporal pairings are constructed by the value-monotone edge-iteration procedure of Section 3.1, and that g_s, g_t are kernel Morse scores of the form in Eq. (2)–Eq. (3) with Lipschitz kernel and aggregation constant $A_{\max}$. If $\mathcal{S}'$ is a data perturbation of $\mathcal{S}$ in the sense of Definition 4 with magnitude $\rho < r_{\mathrm{raw}}(\mathcal{S})$, then $G(\mathcal{S}') = G(\mathcal{S})$, the causal order is preserved ($t_i < t_j \iff t'_i < t'_j$ for all i, j), and the event skeleton is unchanged:* $\mathrm{Skel}(\mathcal{S}') = \mathrm{Skel}(\mathcal{S})$.

In practice, Theorem 1 offer important guarantees that smaller changes in the data are expected to result in smaller changes in the skeletons, thereby ensuring robustness of the topological STTP learning to uncertainties. To illustrate this, we provide a robustness study under various noisy scenarios (see Appendix G).

4.2 Learning Spatio-Temporal Representations

Spatio-Temporal Learner. Positional encoding is a crucial design in the Transformer architecture [46, 16] for making use of the order of the sequence. To effectively utilize temporal information of STPP, for each event time t_i, we encode each timestamp t_i with a sinusoidal embedding defined as $[\boldsymbol{z}_i^t]_j = \cos\left(\frac{t_i}{\xi^{(j-1)/d}}\right)$ for odd j, and $[\boldsymbol{z}_i^t]_j = \sin\left(\frac{t_i}{\xi^{(j-1)/d}}\right)$ for even j, where the frequencies are geometrically scaled by $\xi^{(j-1)/d}$ ($\xi = 10,000$). Then we can obtain a sequence of temporal embedding $\boldsymbol{z}_F^t = (\boldsymbol{z}_1^t, \ldots, \boldsymbol{z}_n^t)$. For the spatial domain, to map them into a compatible latent space, we employ a learnable projection function implemented as an MLP, i.e., $\boldsymbol{z}_i^s = \text{MLP}(s_i)$, and get a corresponding sequence of spatial embeddings $\boldsymbol{z}_F^s = (\boldsymbol{z}_1^s, \ldots, \boldsymbol{z}_n^s)$. This transformation allows the model to capture nonlinear spatial patterns and interactions, producing a sequence of spatial embeddings aligned with the temporal sequence.

Despite obtaining embeddings in a shared latent space, temporal and spatial signals remain fundamentally different in nature. This asymmetry raises an important question: *How can we effectively align and integrate temporal and spatial information such that each modality can inform and refine the other?* To address this question, we fuse the latent embeddings derived from spatial and temporal dimensions to learn bidirectional dependencies between spatial and temporal representations. Specifically, to better integrate these two embeddings, we created two types of attention mechanisms, called spatial attention and temporal attention. Spatial attention uses the spatial embedding as a query and temporal embedding acts as key and value:

$$\boldsymbol{Q}_s^F = \boldsymbol{z}_F^s \boldsymbol{W}_s^Q, \boldsymbol{K}_t^F = \boldsymbol{z}_F^t \boldsymbol{W}_t^K, \boldsymbol{V}_t^F = \boldsymbol{z}_F^t \boldsymbol{W}_t^V, \quad \tilde{\boldsymbol{z}}_F^s = \text{Softmax}(\boldsymbol{Q}_s^F (\boldsymbol{K}_t^F)^\top / \sqrt{d_\ell}) \boldsymbol{V}_t^F, \quad (9)$$

where d_ℓ denotes the dimension of the embedded features. Conversely, the temporal attention uses the temporal embedding as a query and spatial embedding serves as key and value:

$$\boldsymbol{Q}_t^F = \boldsymbol{z}_F^t \boldsymbol{W}_t^Q, \boldsymbol{K}_s^F = \boldsymbol{z}_F^s \boldsymbol{W}_s^K, \boldsymbol{V}_s^F = \boldsymbol{z}_F^s \boldsymbol{W}_s^V, \quad \tilde{\boldsymbol{z}}_F^t = \text{Softmax}(\boldsymbol{Q}_t^F (\boldsymbol{K}_s^F)^\top / \sqrt{d_\ell}) \boldsymbol{V}_s^F. \quad (10)$$

Importantly, these two attention pathways are not redundant, i.e., they capture different perspectives of spatio-temporal interaction where spatial attention emphasizes *"where conditioned on when"*, while temporal attention captures *"when conditioned on where"*. Together, they form a more complete representation of the underlying process. Here we leverage gated fusion to adaptively fuse $\tilde{\boldsymbol{z}}_F^s$ and $\tilde{\boldsymbol{z}}_F^t$ to get the mixed spatio-temporal embedding, i.e., $\tilde{\boldsymbol{z}}_F^{s \leftrightarrow t} = \alpha \tilde{\boldsymbol{z}}_F^s + (1-\alpha) \tilde{\boldsymbol{z}}_F^t$ (where α is a learnable parameter). The resulting fused embedding $\tilde{\boldsymbol{z}}_F^{s \leftrightarrow t}$ provides a unified and context-adaptive representation that captures complex interactions between spatial and temporal dimensions.

Morse-based Spatio-Temporal Learner. At the skeleton level, we extend the above spatio-temporal representation learning framework to capture structured information over skeletal events, i.e., instead of modeling all events, we focus on informative subsets extracted via Morse functions, which provide a compact yet expressive summary of the underlying dynamics. Specifically, we categorize these skeletal events into three types: (i) *spatial Morse-based events*; (ii) *temporal Morse-based events*; and (iii) *spatio-temporal Morse-based events*.

To learn representations for these subsets, we follow a design analogous to the cross-dimensional fusion module described above, while adapting it to the structure of each event type.

For spatial Morse-based events, we employ spatial attention to refine their embeddings by incorporating temporally relevant context: $\tilde{\boldsymbol{z}}_M^s = \text{Softmax}(\boldsymbol{Q}_s^M (\boldsymbol{K}_t^M)^\top / \sqrt{d_\ell}) \boldsymbol{V}_t^M$.

For temporal Morse-based events, we utilize temporal attention to enrich their representations with spatial dependencies: $\tilde{\boldsymbol{z}}_M^t = \text{Softmax}(\boldsymbol{Q}_t^M (\boldsymbol{K}_s^M)^\top / \sqrt{d_\ell}) \boldsymbol{V}_s^M$. For spatio-temporal Morse-based events, which inherently couple both dimensions, we adopt a self-attention mechanism to jointly model their internal spatial-temporal interactions and obtain unified embeddings, i.e., $\tilde{\boldsymbol{z}}_M^{st} = \text{Softmax}(\boldsymbol{Q}_{st}^M (\boldsymbol{K}_{st}^M)^\top / \sqrt{d_\ell}) \boldsymbol{V}_{st}^M$ where $\boldsymbol{Q}_{st}^M, \boldsymbol{K}_{st}^M, \boldsymbol{V}_{st}^M$ are query, key, value matrices derived from spatio-temporal Morse-based events embedding $\boldsymbol{z}_M^{s \leftrightarrow t}$. A gated fusion mechanism further enhances

the spatial and temporal Morse-based outputs by dynamically weighting each modality which is computed as: $\tilde{\boldsymbol{z}}_M^{st} = \beta\tilde{\boldsymbol{z}}_M^{s} + (1-\beta)\tilde{\boldsymbol{z}}_M^{t}$ (where β is a learnable parameter).

Finally, to preserve both the common and complementary information, we apply a concatenation operation: $\bar{\boldsymbol{z}}^{st} = [\tilde{\boldsymbol{z}}_M^{st}, \tilde{\boldsymbol{z}}_F^{s\leftrightarrow t}]$, where $[,]$ denotes the concatenation function and $\bar{\boldsymbol{z}}^{st}$ denotes the final spatio-temporal embedding.

4.3 Learning Geospatial Representations from VLMs

Spatio-temporal representation learning provides a powerful mechanism for integrating temporal and spatial signals in the latent space, yet this integration presupposes that meaningful spatial structure can be extracted from the input. In map-view images, important geospatial patterns such as clustering, dispersion, and directional trends are often implicit rather than explicitly annotated, motivating the need for a model capable of extracting structured and semantically meaningful spatial descriptions directly from such imagery. To this end, we explore whether a VLM can recover structured geographic semantics from map-view imagery and propose an effective prompt (see an example in Fig. 2). Specifically, a prompt $\mathcal{P}_n$ (with n events within a batch) is composed of three parts: (i) The general prompt which defines the model's role as a geospatial expert and specifies the reasoning objectives; (ii) The data-specific description which contextualizes the task; (iii) The instance-level information that encodes the map-view data. For a map-view image $\mathcal{M}_n$, we first use the prompt $\mathcal{P}_n$ and leverage a pre-trained VLM to obtain a joint multimodal representation. Then the VLM embedding is processed by an encoder (here we use a MLP) to transform the VLM embedding into a task-specific representation space as $\boldsymbol{z}^{VLM} = \text{VLM}(\mathcal{M}_n, \mathcal{P}_n, \mathcal{S}_n)$, and $\bar{\boldsymbol{z}}^{VLM} = \text{MLP}(\boldsymbol{z}^{VLM})$, where $\boldsymbol{z}^{VLM}, \bar{\boldsymbol{z}}^{VLM}$ denote map-view image representations learned via a pre-trained VLM and its projected feature respectively.

4.4 Spatio-Temporal Diffusion Model and Morse-Guided Aligner

For each event $\boldsymbol{x}_i = (t_i, s_i)$, following previous studies [19, 54], we apply a K-step Markov chain that progressively perturbs both channels with Gaussian noise under a standard variance-preserving denoising diffusion probabilistic model schedule $\beta_{1:K} \in (0,1)$, factorized across the temporal and spatial dimensions to accommodate their different scales and dynamics. To guide the reverse diffusion toward semantically coherent samples, we condition each denoising step on a unified spatio-temporal embedding, i.e., $\tilde{\boldsymbol{z}}_{i-1} = [\bar{\boldsymbol{z}}_{i-1}^{st}, \bar{\boldsymbol{z}}_{i-1}^{VLM}]$ which compactly encodes the spatio-temporal and VLM-based embeddings. The joint denoising transition decompose is $p_\theta(\boldsymbol{x}_i^{k-1} \mid \boldsymbol{x}_i^k, \bar{\boldsymbol{z}}_{i-1}) = p_\theta(t_i^{k-1} \mid t_i^k, s_i^k, \bar{\boldsymbol{z}}_{i-1})\, p_\theta(s_i^{k-1} \mid t_i^k, s_i^k, \bar{\boldsymbol{z}}_{i-1})$, where p_θ is parameterized by a cross-attentive conditional denoising decoder [49]. At inference time, future events are generated by iteratively applying the learned reverse dynamics.

Multimodal Contrastive Loss. To align learned spatio-temporal representations with geographic semantics, we adopt a symmetric contrastive objective that enforces consistency between the historical-event embedding and the corresponding VLM-based representation. Below, we present the calculation of multimodal contrastive loss on a batch of data:

$$\mathcal{L}_{MCL} = -\frac{1}{2B}\sum_{i=1}^{B}\left[\log\frac{\exp\left(\text{sim}(\bar{\boldsymbol{z}}_i^{st}, \bar{\boldsymbol{z}}_i^{VLM})/\lambda\right)}{\sum_j \exp\left(\text{sim}(\bar{\boldsymbol{z}}_i^{st}, \bar{\boldsymbol{z}}_j^{VLM})/\lambda\right)} + \log\frac{\exp\left(\text{sim}(\bar{\boldsymbol{z}}_i^{VLM}, \bar{\boldsymbol{z}}_i^{st})/\lambda\right)}{\sum_j \exp\left(\text{sim}(\bar{\boldsymbol{z}}_i^{VLM}, \bar{\boldsymbol{z}}_j^{st})/\lambda\right)}\right] \tag{11}$$

where $\text{sim}(\cdot,\cdot)$ denotes cosine similarity, B denotes the batch size, and λ is a temperature.

Morse-Aware Losses. To inject topological inductive bias into the learning objective, we leverage critical structures identified by the spatial and temporal Morse functions. Let $\mathcal{S}_M$ denote the set of spatial critical nodes and $\mathcal{T}_M$ the set of temporal critical nodes where these nodes capture salient structural transitions in space and time, respectively. We encourage the model predictions to align with these topological anchors via two auxiliary losses. Specifically, the spatial loss penalizes the distance between the predicted location and its nearest spatial critical node, while the temporal loss enforces proximity between the predicted inter-arrival time and the nearest temporal critical point. The Morse-aware spatial and temporal losses are defined as follows:

$$\mathcal{L}_{MAL}^{s} = \frac{1}{B}\sum_{i=1}^{B}\min_{\kappa_s \in \mathcal{S}_M}\left\|\hat{s}_i - \kappa_s\right\|_2, \quad \mathcal{L}_{MAL}^{t} = \frac{1}{B}\sum_{i=1}^{B}\min_{\kappa_t \in \mathcal{T}_M}\left|\hat{t}_i - \kappa_t\right| \tag{12}$$

Table 1: Evaluation performance of MCM-DM on 8 datasets.

	Disaster & Epidemic								Urban Dynamics							
	JPN Earthquake		COVID-19		US Tornado		eMAS		Crime		Citibike		Traffic Collision		Twitter	
Model	Spatial↓	Temporal↓	Spatial↓	Temporal↓	Spatial↓	Temporal↓	Spatial↓	Temporal↓	Spatial↓	Temporal↓	Spatial↓	Temporal↓	Spatial↓	Temporal↓	Spatial↓	Temporal↓
Conditional KDE	11.300±0.658	-	0.688±0.047	-	11.872±0.342	-	13.428±1.015	-	0.793±0.004	-	0.718±0.001	-	0.320±0.050	-	0.191±0.009	-
CNF	8.480±0.054	-	0.559±0.000	-	9.564±0.459	-	11.937±0.660	-	0.792±0.003	-	0.720±0.000	-	0.231±0.005	-	0.286±0.007	-
TVCNF	8.110±0.001	-	0.560±0.000	-	9.117±0.285	-	13.284±0.590	-	0.704±0.005	-	0.705±0.000	-	0.209±0.008	-	0.258±0.015	-
Poisson	-	0.631±0.017	-	0.463±0.021	-	0.987±0.003	-	2.478±0.002	-	0.621±0.034	-	0.438±0.001	-	9.183±0.714	-	1.066±0.077
Hawkes	-	0.544±0.010	-	0.672±0.088	-	1.108±0.053	-	2.441±0.0002	-	0.154±0.0016	-	0.534±0.011	-	2.662±0.488	-	1.066±0.077
Self-correcting	-	11.200±0.486	-	2.83±0.141	-	1.137±0.084	-	2.885±0.006	-	0.199±0.015	-	10.700±0.169	-	0.077±0.002	-	16.172±1.095
ODETPP	-	0.479±0.026	-	0.141±0.006	-	0.956±0.015	-	2.468±0.003	-	0.053±0.006	-	0.295±0.043	-	0.034±0.004	-	2.107±2.139
AttNHP	-	0.509±0.005	-	0.232±0.007	-	1.414±0.123	-	2.719±0.089	-	1.176±0.249	-	0.282±0.001	-	1.203±0.221	-	1.222±0.131
RMTPP	-	0.424±0.009	-	1.320±0.024	-	1.096±0.173	-	2.799±0.327	-	0.253±0.159	-	2.070±0.015	-	0.245±0.152	-	0.971±0.147
NHP	-	1.86±0.023	-	2.13±0.100	-	1.044±0.017	-	2.539±0.005	-	0.599±0.035	-	2.360±0.056	-	0.578±0.036	-	0.898±0.031
THP	-	2.44±0.021	-	0.611±0.008	-	1.812±0.734	-	2.911±0.372	-	1.391±0.819	-	1.460±0.009	-	1.489±0.871	-	1.750±0.743
SAHP	-	0.409±0.002	-	0.184±0.024	-	1.319±0.200	-	2.754±0.103	-	0.994±0.224	-	0.279±0.002	-	1.080±0.304	-	1.136±0.236
NJSDE	9.98±0.024	0.465±0.009	0.641±0.009	0.137±0.001	9.731±0.219	1.072±0.046	12.108±0.273	2.463±0.081	0.067±0.002	0.038±0.0001	0.707±0.001	0.264±0.005	0.142±0.006	0.014±0.001	0.173±0.009	0.912±0.031
NSTPP	8.11±0.000	0.547±0.010	0.560±0.000	0.145±0.002	9.905±0.231	1.096±0.053	12.264±0.295	2.538±0.087	0.081±0.003	0.041±0.002	0.705±0.000	0.355±0.013	0.169±0.007	0.018±0.002	0.185±0.008	0.829±0.034
DeepSTPP	9.20±0.000	**0.341±0.000**	0.687±0.000	0.197±0.000	10.978±0.243	1.115±0.058	13.341±0.307	2.789±0.092	0.075±0.004	0.044±0.002	0.044±0.000	0.234±0.000	0.156±0.008	0.062±0.002	0.193±0.011	0.846±0.037
DSTPP	6.77±0.193	0.375±0.001	0.546±0.013	0.114±0.001	10.087±0.326	1.110±0.049	14.079±0.482	2.610±0.119	0.080±0.003	0.041±0.0001	0.044±0.0006	0.296±0.015	0.179±0.0004	0.012±0.0002	0.099±0.001	1.885±0.035
TEN-DM	6.649±0.041	0.371±0.000	0.391±0.001	0.1035±0.0010	7.214±0.187	0.983±0.0089	4.232±0.041	0.581±0.130	0.072±0.0003	0.051±0.0005	0.043±0.0022	0.256±0.0037	0.062±0.0003	0.012±0.0001	0.096±0.007	0.883±0.028
MCM-DM (Ours)	**6.3890±0.1323**	0.3751±0.0002	**0.187±0.0001**	**0.095±0.0000**	**4.134±0.1886**	**0.960±0.0081**	**2.555±0.0310**	**0.304±0.1070**	**0.039±0.0002**	**0.029±0.0001**	**0.014±0.0001**	**0.206±0.0025**	**0.026±0.0001**	**0.009±0.0000**	**0.041±0.0014**	**0.737±0.0205**

where $\hat{s}_i$ is the predicted clean location at a randomly sampled diffusion step and $\hat{t}_i$ is the predicted inter-arrival time. Finally, the overall loss function is defined as the sum of all 4 components, i.e., $\mathcal{L}_{total} = \mathcal{L}_{Diff} + \eta_1 \mathcal{L}_{MCL} + \eta_2 \mathcal{L}^s_{MAL} + \eta_3 \mathcal{L}^t_{MAL}$, where $\mathcal{L}_{Diff}$ is the standard denoising diffusion loss (here we use L_2 loss), and η_1, η_2, η_3 are hyperparameters.

5 Experiments

Datasets. In our experiments, we use 8 real-world datasets: JPN Earthquake, COVID-19, US Tornado, eMAS, Crime, Citibike, Traffic Collision, and Twitter. More details of data resources and train/validation/test split ratio are in Appendix H. **Baselines and Implementation Details.** To examine the effectiveness of our method across diverse data, we compare MCM-DM with 17 baseline models that cover a range of architectures. We run our experiments on 4 NVIDIA RTX A5000 GPU cards with 24GB memory. Evaluation metrics include spatial Euclidean distance and temporal root mean square error (RMSE), averaged over 3 runs with different random seeds. The best performance is in **red**, and the second-best is underlined. More details can be found in Appendix H.

Main Results. Table 1 presents spatial and temporal forecasting performance of the proposed MCM-DM compared with baselines. MCM-DM achieves outstanding performance and outperforms baselines on *7 out of 8* datasets on temporal dimension and *all* datasets on spatial dimensions. Specifically, on disaster and epidemic datasets (see Table 1 left), our MCM-DM brings up to 87.07% average gain relative to runner-ups.

For Japan Earthquake data's temporal dimension, as presented in Fig. 3, Appendix F, we explain that data is nearly uniform with weak temporal dependence, thus the advantage of structured spatio-temporal modeling is diminished. On urban dynamics datasets (see Table 1 right), MCM-DM has *statistically significant* improvements compared to conditional spatial and temporal baselines and neural network-based spatio-temporal baselines (*average p-values are* < 0.001 *and* < 0.05 for spatial and temporal dimensions respectively), with notably highly significant gains for Citibike and Traffic Collision.

The superior performance of MCM-DM can be attributed to several key architectural innovations: (i) Morse-theoretic approaches highlight topologically critical events and suppresses noise; (ii) VLM-based geographic encoder that imparts semantic context beyond raw coordinates; and (iii) Morse-guided aligner aligns spatial, temporal, and semantic representations in a unified latent space. The computational complexity and running time comparison (see Table 11) can be found in Appendix I. It can be seen that MCM-DM could **decrease the running time significantly**, demonstrating its great efficiency for spatio-temporal forecasting.

Table 2: Ablation studies.

	COVID-19		Traffic Collision		Citibike		Twitter	
Model	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓
MCM-DM	**0.1866±0.0001**	**0.0953±0.0000**	**0.0264±0.0001**	**0.0088±0.0000**	**0.0145±0.0001**	**0.2061±0.0025**	**0.0412±0.0014**	**0.7373±0.0205**
W/o VLM	0.4283±0.0078	0.0988±0.0015	0.1263±0.0006	0.0098±0.0002	0.0325 ±0.0023	0.2278±0.0178	0.0852±0.0016	0.865±0.042
W/o STL	0.4949±0.0050	0.1058±0.0029	0.1532±0.0013	0.0097±0.0001	0.0365±0.0002	0.2382±0.0018	0.0853±0.0014	0.8905±0.0737
W/o MSTL	0.2173±0.0015	0.1077±0.0014	0.0297±0.0001	0.0097±0.0001	0.0165±0.0001	0.2411±0.0080	0.0563±0.0010	0.8071±0.0054
W/o MAL	0.2210±0.0030	0.1109±0.0029	0.0299±0.0002	0.0097±0.0001	0.0165±0.0002	0.2380±0.0046	0.0674±0.0135	0.971±0.270

Ablation Studies, Sensitivity Analysis, & Robustness Analysis. As our MCM-DM incorporates 4 components, i.e., VLM-based encoder (VLM), spatial-temporal learner (STL), Morse-based

Table 3: Sensitivity Analysis.

# of Neighbors (k)	Crime		US Tornado		Citibike		Twitter	
	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓
3	0.0438±0.0006	0.0320±0.0005	4.1778±0.1669	**0.9516±0.0054**	0.0165±0.0000	0.2392±0.0055	**0.0441±0.0004**	0.7718±0.0118
4	0.0434±0.0009	0.0312±0.0005	4.3978±0.2843	0.9645±0.0037	0.0165±0.0001	0.2404±0.0125	0.0445±0.0007	0.7743±0.0330
5	**0.0434±0.0018**	**0.0307±0.0008**	4.2245±0.1847	0.9669±0.0031	**0.0165±0.0001**	**0.2379±0.0046**	0.0449±0.0015	**0.7654±0.0099**
6	0.0438±0.0013	0.0314±0.0007	**4.1338±0.1886**	0.9600±0.0081	0.0165±0.0001	0.2391±0.0047	0.0444±0.0005	0.7769±0.0094

spatio-temporal learner (MSTL), and Morse-guided aligner (MAL), we assess their effectiveness by removing the corresponding component from MCM-DM. Table 2 presents the results of these ablations on 4 datasets. We find that removing: (i) the VLM-based encoder leads to consistent degradation across all scenarios; (ii) the STL results in noticeable performance drops in both spatial and temporal metrics; (iii) the MSTL consistently degrades performance, with more pronounced drops in the temporal metrics; and (iv) the Morse-guided aligner leads to noticeable instability and performance deterioration. Table 3 evaluates the MCM-DM sensitivity to the number of neighbors k in the kNN graph. Overall, the performance remains stable across different choices of k. Moderate values (e.g., $k = 4, 5$) consistently yield the best or near-best results across datasets which suggests a good balance between capturing local structure and avoiding noisy connections. We can also recommend to use cross-validation for selecting optimal k. Furthermore, loss weights analysis of Section 4.4 (see Table 6) and robustness analysis (see Table 9) are provided in Appendix G.

6 Conclusion

We presented MCM-DM, a novel framework that integrates discrete Morse theory with diffusion-based modeling to advance STPP learning. By introducing a Morse-aware representation pipeline, MCM-DM captures not only event dynamics but also the underlying structural organization of event sequences, and Morse-guided aligner further bridges modality gaps by anchoring cross-modal interactions to topologically critical events which enables more faithful and stable integration of semantic geographic context. Extensive experiments on 8 real-world datasets demonstrate that MCM-DM consistently achieves superior performance in both spatial and temporal forecasting, and maintains computational efficiency.

In the future we plan to advance the proposed ideas on the Morse-aware representations to topological uncertainty quantification in spatio-temporal point processes, by harnessing the resampling ideas similar to bootstrap over zigzag persistent homology [5]. The other interesting direction to investigate is a holistic integration of prior expert knowledge to the selection of a suitable score function.

Acknowledgement

This work has been supported in part by NSF DMS-2533984/2533985 and OAC-2530469/253047, by the U.S. Department of Energy (DOE), Office of Science, Advanced Scientific Computing Research (ASCR) program under the Scientific Discovery through Advanced Computing (SciDAC) Institute "LEADS: LEarning-Accelerated Domain Science". This research was conducted in part while Y.R.G. visited the Okinawa Institute of Science and Technology (OIST) through the Theoretical Sciences Visiting Program (TSVP). The authors are also very thankful to the Area Chair and the three reviewers for their thought provoking feedback and many motivating questions opening a path for future directions to address.

References

[1] Sunil Arya, David M. Mount, Nathan S. Netanyahu, Ruth Silverman, and Angela Y. Wu. An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *Journal of the ACM*, 45(6):891–923, 1998.

[2] Biswadeep Chakraborty, Hemant Kumawat, Beomseok Kang, and Saibal Mukhopadhyay. Dynamic graph structure estimation for learning multivariate point process using spiking neural networks. *arXiv preprint arXiv:2504.01246*, 2025.

[3] Ricky T Q Chen, Brandon Amos, and Maximilian Nickel. Neural spatio-temporal point processes. In *ICLR*, 2021.

[4] Ricky T Q Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *NeurIPS*, 2018.

[5] Yuzhou Chen and Yulia Gel. Topological zigzag spaghetti for diffusion-based generation and prediction on graphs. In *ICLR*, 2025.

[6] Xiuyuan Cheng, Zheng Dong, and Yao Xie. Deep spatiotemporal point processes: Advances and new directions. *Annual Review of Statistics and Its Application*, 13, 2025.

[7] Wen-Hao Chiang, Xueying Liu, and George Mohler. Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, 38(2):505–520, 2022.

[8] David R Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society: Series B*, 17(2):129–157, 1955.

[9] Daryl J Daley and David Vere-Jones. *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*. Springer, 2003.

[10] Peter J Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. CRC Press, 2013.

[11] Zheng Dong, Xiyuan Cheng, and Yao Xie. Spatio-temporal point processes with deep non-stationary kernels. In *ICLR*, 2023.

[12] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. Recurrent marked temporal point processes: Embedding event history to vector. In *ACM SIGKDD*, 2016.

[13] Robin Forman. Morse theory for cell complexes. *Advances in Mathematics*, 134(1):90–145, 1998.

[14] Robin Forman. A user's guide to discrete morse theory. *Séminaire Lotharingien de Combinatoire*, 48:B48c, 2002.

[15] Jerome H. Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3(3):209–226, 1977.

[16] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. Convolutional sequence to sequence learning. In *ICML*, 2017.

[17] Alan G Hawkes. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 33(3):438–443, 1971.

[18] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90, 1971.

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

[20] Valerie Isham and Mark Westcott. A self-correcting point process. *Stochastic processes and their applications*, 8(3):335–347, 1979.

[21] Junteng Jia and Austin R. Benson. Neural jump stochastic differential equations. In *NeurIPS Reproducibility Challenge*, 2019.

[22] Ming Jin, Huan Yee Koh, Qingsong Wen, Daniele Zambon, Cesare Alippi, Geoffrey I Webb, Irwin King, and Shirui Pan. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE transactions on pattern analysis and machine intelligence*, 46(12):10466–10485, 2024.

[23] Michael Joswig and Marc E. Pfetsch. Computing optimal Morse matchings. *SIAM Journal on Discrete Mathematics*, 20(1):11–25, 2006.

[24] Harish Kannan, Emil Saucan, Indrava Roy, and Areejit Samal. Persistent homology of unweighted complex networks via discrete Morse theory. *Scientific Reports*, 9:13817, 2019.

[25] Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. In *AAAI*, 2025.

[26] Fan Liu, Delong Chen, Zhangqingyun Guan, Xiaocong Zhou, Jiale Zhu, Qiaolin Ye, Liyong Fu, and Jun Zhou. Remoteclip: A vision language foundation model for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[27] Yuxin Liu, Kaiming Wang, Chenguang Yang, Yulia Gel, and Yuzhou Chen. TEN-DM: Topology-enhanced diffusion model for spatio-temporal event prediction. In *ICLR*, 2026.

[28] David Lüdke, Marin Biloš, Oleksandr Shchur, Marten Lienen, and Stephan Günnemann. Add and thin: Diffusion for temporal point processes. In *NeurIPS*, 2023.

[29] Antonino Marciano, Emanuele Zappala, Tommaso Torda, Matteo Lulli, Stefano Giagu, Chris Fields, Deen Chen, and Filippo Fabrocini. Deep neural networks as the semi-classical limit of topological quantum neural networks: The problem of generalisation. *arXiv preprint arXiv:2210.13741*, 2022.

[30] Hongyuan Mei and Jason Eisner. The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process. In *NeurIPS*, 2017.

[31] Hongyuan Mei, Chenghao Yang, and Jason Eisner. Transformer embeddings of irregularly spaced events and their participants. In *ICLR*, 2021.

[32] John Milnor. *Lectures on the h-cobordism theorem*, volume 2258. Princeton university press, 2025.

[33] George O. Mohler, Martin B. Short, P. Jeffrey Brantingham, Frederic P. Schoenberg, and George E. Tita. Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106(493):100–108, 2011.

[34] Sumantrak Mukherjee, Mouad Elhamdi, George Mohler, David A Selby, Yao Xie, Sebastian Vollmer, and Gerrit Grossmann. Neural spatiotemporal point processes: Trends and challenges. *TMLR*, 2025.

[35] Jerzy Neyman and Elizabeth L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(1):1–29, 1958.

[36] Yosihiko Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988.

[37] Yosihiko Ogata. Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402, 1998.

[38] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA, 2006.

[39] Murray Rosenblatt. Conditional probability density and regression estimators. *Multivariate analysis II*, 25:31, 1969.

[40] Salva Rühling Cachay, Bo Zhao, Hailey Joren, and Rose Yu. Dyffusion: A dynamics-informed diffusion model for spatiotemporal forecasting. In *NeurIPS*, 2023.

[41] Oleksandr Shchur, Marin Biloš, and Stephan Günnemann. Intensity-free learning of temporal point processes. In *ICLR*, 2020.

[42] Oleksandr Shchur, Ali Caner Türkmen, Tim Januschowski, and Stephan Günnemann. Neural temporal point processes: A review. *IJCAI Survey Track*, 2021.

[43] Stephen Smale. On gradient dynamical systems. *Annals of Mathematics*, 74(1):199–206, 1961.

[44] Yujee Song, Donghyun Lee, Rui Meng, and Won Hwa Kim. Decoupled marked temporal point process using neural ordinary differential equations. In *ICLR*, 2024.

[45] Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer, New York, 1999.

[46] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

[47] Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. In *NeurIPS*, 2023.

[48] Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, et al. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*, 2024.

[49] Yu Wang, Zhiwei Liu, Liangwei Yang, and Philip S Yu. Conditional denoising diffusion for sequential recommendation. In *PAKDD*, 2024.

[50] Chengyuan Wu, Shiquan Ren, Jie Wu, and Kelin Xia. Discrete morse theory for weighted simplicial complexes. *Topology and its Applications*, 270:107038, February 2020.

[51] Jiue-An Yang, Yuzhou Chen, Calvin Tribby, Loretta Erhunmwunsee, Caroline Thompson, Tarik Benmarhnia, Hugo Kyo Lee, Marta Jankowska, and Yulia R Gel. Understanding the impact of environmental contexts on lung cancer with simplicial representation learning and remote sensing. In *IEEE BigData*, pages 3315–3324, 2025.

[52] Jiue-An Yang, Yuzhou Chen, Calvin P Tribby, Huikyo Lee, Loretta Erhunmwunsee, Tarik Benmarhnia, Caroline A Thompson, Yulia R Gel, and Marta M Jankowska. Llm-based multi-agent system and simplicial self-supervised learning model for regional cancer prevalence estimation using satellite imagery. In *SIGSPATIAL*, pages 1182–1185, 2025.

[53] Yiyuan Yang, Ming Jin, Haomin Wen, Chaoli Zhang, Yuxuan Liang, Lintao Ma, Yi Wang, Chenghao Liu, Bin Yang, Zenglin Xu, Shirui Pan, and Qingsong Wen. A survey on diffusion models for time series and spatio-temporal data. *ACM Computing Surveys*, 58(8):1–39, 2026.

[54] Yuan Yuan, Jingtao Ding, Chenyang Shao, Depeng Jin, and Yong Li. Spatio-temporal diffusion point processes. In *KDD*, 2023.

[55] Qiang Zhang, Aldo Lipani, Omer Kirnap, and Emine Yilmaz. Self-attentive Hawkes process. In *ICML*, 2020.

[56] Shuai Zhang, Chuan Zhou, Yang Aron Liu, Peng Zhang, Xixun Lin, and Zhi-Ming Ma. Neural jump-diffusion temporal point processes. In *ICML*, 2024.

[57] Yuanheng Zhang, Chenguang Yang, Nicholas LaHaye, Huikyo Lee, Yulia R. Gel, and Yuzhou Chen. Uncertainty-aware multi-modality spatio-temporal wildfire and smoke prediction. In *SIGSPATIAL*, 2026.

[58] Wenjun Zhao and Esteban G Tabak. Adaptive kernel conditional density estimation. *Information and Inference: A Journal of the IMA*, 14(1):iaae037, 2025.

[59] Zihao Zhou, Xingyi Yang, Ryan Rossi, Handong Zhao, and Rose Yu. Neural point process for learning spatiotemporal event dynamics. In *L4DC*, 2022.

[60] Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. Transformer Hawkes process. In *ICML*, 2020.

A Technical appendices and supplementary material

Table 4: Closed-form Matérn kernels at half-integer smoothness values.

ν	$K_\nu(r)$	Common name
$1/2$	e^{-r}	Exponential (Laplace)
$3/2$	$\left(1+\sqrt{3}\,r\right)e^{-\sqrt{3}\,r}$	Matérn $3/2$
$5/2$	$\left(1+\sqrt{5}\,r+\frac{5}{3}\,r^2\right)e^{-\sqrt{5}\,r}$	Matérn $5/2$
∞	$e^{-r^2/2}$	Gaussian (squared-exponential)

Comparison. Table 5 summarises smoothness, support, and decision-gap Lipschitz constants across the Matérn family and the two non-Matérn baselines. Within the Matérn family, $\nu = 5/2$ is the safest default for general STPPs; $\nu = 1/2$ (exponential) is preferred when the STPP exhibits documented Hawkes-style self-excitation; $\nu = \infty$ (Gaussian) is preferred for very smooth phenomena. Outside the family, Epanechnikov is useful when only local recurrence is of interest; Tophat is the degree-based baseline.

Table 5: Comparison of Matérn-family choices and non-Matérn baselines. $\|K'\|_\infty$ is the maximum slope of the kernel (numerical for Matérn $3/2, 5/2$). L is the data-direction decision-gap Lipschitz constant from (15); the factor $4k/\sigma$ is suppressed for clarity. Structural emphasis describes which STPP feature each kernel highlights as critical.

Family	Kernel (ν or label)	$\|K'\|_\infty$	Smoothness	Structural emphasis
Matérn	$K_{1/2}$ (exponential)	1.000	C^0 at 0	Hawkes self-excitation; heavy-tail
Matérn	$K_{3/2}$	0.638	C^1	Once-differentiable; balanced
Matérn	$K_{5/2}$	0.625	C^2	Twice-differentiable; default
Matérn	K_∞ (Gaussian)	0.607	C^∞	Smooth long-range coupling

B Background: Discrete Morse Theory

B.1 Cobordism as a Trajectory of Topological Change

Rather than studying spaces in isolation, *cobordism theory* studies the *process by which one space transforms into another*. A cobordism is a higher-dimensional object whose boundary consists of two lower-dimensional spaces, encoding an entire topology-changing trajectory between them.

Intuitively, a cobordism can be viewed as a long-exposure photograph of a person walking across a room. Rather than capturing a single frozen moment, the image records the entire journey as a blurred light trail from start to finish. A cobordism encodes an object *changing through time*: not a static snapshot, but the full trajectory itself. Where classical geometry studies objects at fixed moments, cobordisms study the transformation—the morphism—connecting one state to another. Crucially, not every point along this trajectory is equally significant: certain moments, such as changes in direction or pauses before motion, mark meaningful transitions.

Definition 8 (Cobordism)**.** Two closed $(d-1)$-manifolds M and N are *cobordant* if there exists a d-manifold W such that:

$$\partial W = M \sqcup N.$$

B.2 Morse Theory as the Identification of Topological Events

Morse theory provides the operative bridge between cobordisms and dynamics. A Morse function on a cobordism decomposes the transformation into elementary topological events occurring at critical points. Between critical values, the topology is invariant; all qualitative change is localized at critical events. Thus, a Morse function induces a canonical *coarse-to-fine* description of structural evolution: topology first, geometry later.

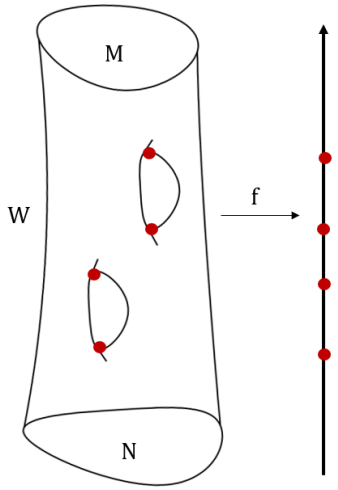


Figure 1: **Cobordism depiction.** The tube W represents a nontrivial cobordism W between M and N. As we sweep through the Morse function f, the topology of level sets evolves: the first index-1 saddle splits this circle into two disjoint loops, the second saddle re-merges into one and the cycle is repeated again when the second hole is reached. This sequence of topological transitions (birth, splitting, merging, death) demonstrates how a cobordism can interpolate between boundary conditions while traversing rich intermediate topology.

Definition 9 (Critical point (smooth))**.** Let M be a smooth n-dimensional manifold and let $f : M \to \mathbb{R}$ be a smooth function. A point $p \in M$ is a *critical point* of f if the differential of f vanishes at p, i.e., $df_p = 0$. Equivalently, in any local coordinate chart $(x_1, \ldots, x_n)$ around p,

$$\frac{\partial f}{\partial x_i}(p) = 0 \quad \text{for all } i = 1, \ldots, n.$$

A point p is a *regular point* if $df_p \neq 0$.

Definition 10 (Morse function (smooth))**.** A smooth function $f : M \to \mathbb{R}$ is a *Morse function* if all of its critical points are non-degenerate: for every critical point p, the Hessian $\mathrm{Hess}_p(f)$ is non-singular.

Non-degeneracy implies that critical points are isolated and finite in number for compact M.

C Discrete Morse Theory

We now turn to the discrete analogue of smooth Morse theory introduced by Forman [13, 14], which replaces smooth manifolds with cell complexes. This setting is essential for graphs, since the clique complex of a (weighted or unweighted) graph generally admits no smooth structure.

C.1 Graphs and Clique Complexes

Definition 11 (Clique complex of a graph)**.** Let $G = (V, E, \boldsymbol{\mathcal{X}})$ be a finite simple graph. The *clique complex* $K = \mathrm{Cl}(G)$ is the abstract simplicial complex whose simplices are the vertex sets of cliques in G: a p-simplex is a set of $p+1$ mutually adjacent nodes. Hence nodes, edges, and triangles correspond to 0-, 1-, and 2-simplices, respectively. The complex is downward closed: if $\alpha \in K$, then every non-empty subset of α lies in K. We write $\dim(\alpha) = |\alpha| - 1$, $\dim(K) = \max_{\alpha \in K} \dim(\alpha)$, and $\gamma < \alpha$ to denote that γ is a proper face of α.

C.2 Discrete Morse Functions and Critical Simplices

Let K be a simplicial complex and let $f : K \to \mathbb{R}$ assign a real value to each simplex. For a p-simplex $\alpha \in K^p$, define

$$U_f(\alpha) = \{\beta \in K^{p+1} \mid \alpha < \beta,\ f(\beta) \le f(\alpha)\}, \tag{13}$$

$$V_f(\alpha) = \{\gamma \in K^{p-1} \mid \gamma < \alpha,\ f(\gamma) \ge f(\alpha)\}. \tag{14}$$

Intuitively, $U_f(\alpha)$ contains cofaces of α with non-increasing function value, while $V_f(\alpha)$ contains faces of α with non-decreasing value.

Definition 12 (Discrete Morse function)**.** The function $f : K \to \mathbb{R}$ is a *discrete Morse function* if for every simplex $\alpha \in K$,

$$|U_f(\alpha)| \le 1 \quad \text{and} \quad |V_f(\alpha)| \le 1.$$

Definition 13 (Critical simplex)**.** A simplex $\alpha \in K$ is *critical* if

$$|U_f(\alpha)| = |V_f(\alpha)| = 0.$$

Equivalently, all cofaces of α have strictly larger function values and all faces of α have strictly smaller function values. We write $\mathcal{C}(f)$ for the set of all critical simplices of f, and $\mathcal{C}_p(f) = \{\alpha \in K^p : \alpha \text{ is critical}\}$ for the critical p-simplices.

Critical simplices play the role of critical points in smooth Morse theory, in that they encode the irreducible topological features of K and form the basis for Morse reductions and topology-aware filtrations used in persistent homology.

C.3 Discrete Gradient Vector Fields

Definition 14 (Discrete gradient vector field)**.** A *discrete gradient vector field* $\mathcal{V}$ on K is a collection of pairs $(\alpha^{(p)}, \beta^{(p+1)})$ with $\alpha < \beta$, such that each cell appears in at most one pair. We require $\mathcal{V}$ to be *acyclic*: there are no nontrivial closed $\mathcal{V}$-paths.

A simplex is critical (Definition 13) if and only if it is unpaired in $\mathcal{V}$, i.e., it participates in no gradient arrow.

C.4 The Directed Hasse Diagram and $\mathcal{V}$-Paths

Definition 15 (Directed Hasse diagram)**.** Let K be a finite regular CW complex and let $\mathcal{V}$ be a discrete gradient vector field on K. Define the *directed Hasse diagram* $G_{\mathcal{V}}$ on the cells of K by orienting each incidence $\alpha^{(p)} < \beta^{(p+1)}$ as

$$\beta \mapsto \alpha \quad \text{unless} \quad (\alpha, \beta) \in \mathcal{V}, \text{ in which case } \alpha \mapsto \beta.$$

Definition 16 ($\mathcal{V}$-path)**.** A *$\mathcal{V}$-path* is a directed path in $G_{\mathcal{V}}$ alternating dimensions:

$$\alpha_0^{(p)} < \beta_0^{(p+1)} > \alpha_1^{(p)} < \beta_1^{(p+1)} > \cdots,$$

where $(\alpha_i, \beta_i) \in \mathcal{V}$ for each i and α_{i+1} is a codimension-1 face of β_i with $\alpha_{i+1} \neq \alpha_i$.

C.5 Descending Manifolds and Descending Paths

Definition 17 (Descending manifold / descending set)**.** For a critical p-cell $\sigma \in \mathcal{C}(f)$, its *descending manifold* (or descending set) is

$$\mathcal{D}_{\mathcal{V}}(\sigma) := \{\tau \in K \mid \exists \text{ a } \mathcal{V}\text{-path from } \sigma \text{ reaching } \tau\}.$$

Equivalently, $\tau \in \mathcal{D}_{\mathcal{V}}(\sigma)$ if τ is reachable from σ by following the directed incidences of $G_{\mathcal{V}}$. Intuitively, $\mathcal{D}_{\mathcal{V}}(\sigma)$ describes the region of the complex whose combinatorial flow terminates at σ.

Rather than retaining the full descending manifold of each critical cell, which may be large and unstable under benign perturbations, we associate to each critical cell a single representative path:

Definition 18 (Descending path)**.** For a critical cell $\alpha \in \mathcal{C}(f)$, a *descending path* $\pi(\alpha)$ is a deterministic gradient-respecting walk in $G_{\mathcal{V}}$ contained in $\mathcal{D}_{\mathcal{V}}(\alpha)$:

$$\pi(\alpha) = (\alpha = \sigma_0 \to \sigma_1 \to \cdots \to \sigma_m),$$

where each step follows one gradient arrow (or the induced directed adjacency) toward strictly lower f-value, terminating at a critical cell.

In the 1-skeleton setting, the discrete gradient vector field $\mathcal{V}$ induces a unique outgoing move at each nonterminal simplex, so each critical edge deterministically follows a finite $\mathcal{V}$-path to a terminal critical cell, typically a vertex. The walk alternates between (i) "up" moves along the unique gradient pairing (when present) and (ii) "down" moves to an allowed face distinct from the paired face, enforcing strict decrease in f on each down move.

C.6 The Morse Skeleton

Definition 19 (Morse skeleton)**.** Let G be a graph with clique complex $\mathrm{Cl}(G)$ and discrete Morse function f, with associated set of critical cells $\mathcal{C}(f)$. The *Morse skeleton* of G is the finite set

$$\mathcal{S}(G) := (\mathcal{C}, \mathcal{D}) = \big(\mathcal{C}(f),\ \{(\alpha, \pi(\alpha)) : \alpha \in \mathcal{C}(f)\}\big),$$

where $\pi(\alpha)$ is the descending path associated to each critical cell (Definition 18). The skeleton encodes both the locations of topological events (critical cells) and a minimal canonical description of how combinatorial flow propagates from them.

C.7 Filtration Compatibility

Definition 20 (Compatibility notions)**.** Let K be a finite cell complex, $f : K \to \mathbb{R}$ a discrete Morse function, and $\mathcal{V}$ a discrete gradient vector field.

1. **f-admissible (induced-filtration compatible).** $\mathcal{V}$ is *f-admissible* if for every matched pair $(\sigma^{(p)}, \tau^{(p+1)}) \in \mathcal{V}$,
$$f(\tau) \le f(\sigma).$$
Equivalently, if $F_t := \{\alpha \in K : f(\alpha) \le t\}$ is the sublevel-set filtration, then $\sigma \in F_t$ implies its paired coface (if any) also lies in F_t.

2. **Strict filtered matching.** Given an externally specified filtration $F : K_0 \subset \cdots \subset K_T$ with entry time $t_F(\alpha)$, $\mathcal{V}$ is *entry-time compatible* if every matched pair satisfies $t_F(\sigma) = t_F(\tau)$.

C.8 Forman's Morse Inequalities and Topology Preservation

Let c_p denote the number of critical p-cells of a discrete Morse function f on K. Forman's inequalities state that for all p,

$$c_p \ge \beta_p(K),$$

and the alternating sum equality

$$\sum_p (-1)^p c_p = \chi(K)$$

holds, where $\chi(K)$ is the Euler characteristic. Moreover, K admits a homotopy-equivalent CW complex with exactly one p-cell for each critical p-cell, so critical cells fully determine the homotopy type of K.

C.9 Forestedness, H_0-Faithfulness, and the Balance Identity

The semantic interpretation of the four-way partition rests on each spatial and temporal pairing inducing a spanning forest, so that critical edges count cycle rank exactly.

Definition 21 (Forested pairing)**.** V is *forested* if the paired-edge subgraph (V, E^{pair}) is acyclic as an undirected graph.

Proposition 1 (Exact balance identity under H_0-faithfulness and forestedness)**.** *Suppose the discrete gradient vector fields induced by g_s and g_t are both H_0-faithful and forested. Then both paired-edge subgraphs (V, E_s^{pair}) and (V, E_t^{pair}) are spanning forests of G, and $|E_s^{\mathrm{crit}}| = |E_t^{\mathrm{crit}}| = \beta_1(G)$. Consequently $|E_{cc}| + |E_{cp}| = |E_{cc}| + |E_{pc}| = \beta_1(G)$, so the four quadrant sizes satisfy the balance identity $|E_{cp}| = |E_{pc}|$.*

The proof follows from strict g-monotonicity along any V-path, which a closed loop would contradict [13].

D Types of Critical Edges

Under H_0-faithfulness for g_s and g_t, both E_s^{pair} and E_t^{pair} are spanning forests and $|E_s^{\mathrm{crit}}| = |E_t^{\mathrm{crit}}| = \beta_1(G)$. The critical edge *counts* are identical, but what differs is *which* edges are critical, as the two functions select different sets of β_1 cycle-closing edges. The intersection of the two decompositions yields a four-way partition $E = E_{cc} \sqcup E_{cp} \sqcup E_{pc} \sqcup E_{pp}$ where $E_{cc} = E_s^{\mathrm{crit}} \cap E_t^{\mathrm{crit}}$ (both-critical), $E_{cp} = E_s^{\mathrm{crit}} \cap E_t^{\mathrm{pair}}$ (spatial-cycle, temporal-tree), $E_{pc} = E_s^{\mathrm{pair}} \cap E_t^{\mathrm{crit}}$ (spatial-tree, temporal-cycle), and $E_{pp} = E_s^{\mathrm{pair}} \cap E_t^{\mathrm{pair}}$ (both-paired).

Quadrant	STPP interpretation
E_{cc}	*Spatio-temporal recurrence.* Activity returning to the same place at a causally linked time (aftershock on same fault, repeat crime at same location, seasonal epidemic flare-up).
E_{cp}	*Persistent hotspot.* Spatially recurrent but temporally regular—a chronically active location with steady event rate (endemic disease, permanent seismic zone).
E_{pc}	*Travelling wave.* Spatially sequential but temporally recurrent—activity migrating through space with bursty or periodic temporal pattern (aftershock migration along a fault, spreading outbreak).
E_{pp}	*Backbone.* Tree-like under both decompositions—the expected, density-following transitions forming the bulk of the event cloud.

E Proofs

E.1 Decision Margin for the Four-Way Partition

The greedy Forman pairing applied to either g_s or g_t reduces to a finite collection of strict comparisons between Morse values at adjacent nodes. Namely, if no comparison flips its sign, neither decomposition $E = E_s^{\mathrm{crit}} \sqcup E_s^{\mathrm{pair}}$ nor $E = E_t^{\mathrm{crit}} \sqcup E_t^{\mathrm{pair}}$ changes, and consequently neither does their intersection $\mathrm{Skel}(\mathcal{S})$.

Definition 22 (Decision gap)**.** For each adjacent node pair $\{x_i, x_j\}$ in G, define the spatial and temporal decision gaps

$$\delta_{ij}^{(s)}(\mathcal{S}) \;:=\; g_s(x_i) - g_s(x_j), \qquad \delta_{ij}^{(t)}(\mathcal{S}) \;:=\; g_t(x_i) - g_t(x_j).$$

Let $J = \{(i,j) : \{x_i, x_j\} \in E\}$ index the comparisons performed by greedy pairing.

Definition 23 (Skeleton decision margin)**.** Assuming $\delta_{ij}^{(s)} \neq 0$ and $\delta_{ij}^{(t)} \neq 0$ for all $(i,j) \in J$ (the no-ties condition), the *skeleton decision margin* of $\mathcal{S}$ is

$$m(\mathcal{S}) \;:=\; \min_{(i,j)\in J} \min\Big\{ \big|\delta_{ij}^{(s)}(\mathcal{S})\big|,\; \big|\delta_{ij}^{(t)}(\mathcal{S})\big| \Big\}.$$

The no-ties condition is generic and is always achievable by the infinitesimal jitter $\varepsilon_i^{(s)}, \varepsilon_i^{(t)}$ already included in the Morse-function definitions.

Definition 24 (Decision-gap Lipschitz constants)**.** Let $L_s, L_t \geq 0$ be the smallest constants such that, for any two STPP realizations $\mathcal{S}, \mathcal{S}'$ on the same node set differing by a data perturbation of magnitude ρ,

$$\left|\delta_{ij}^{(s)}(\mathcal{S}) - \delta_{ij}^{(s)}(\mathcal{S}')\right| \leq L_s\,\rho, \quad \left|\delta_{ij}^{(t)}(\mathcal{S}) - \delta_{ij}^{(t)}(\mathcal{S}')\right| \leq L_t\,\rho,$$

for every $(i, j) \in J$.

Explicit bounds for L_s, L_t in terms of the kernel slope $\|K'\|_\infty$ and the aggregation constant $A_{\max}$ are established in Proposition 4.

Proposition 2 (Balance identity for the four-way Morse partition)**.** *Let $G = (V, E)$ be a finite graph, viewed as a one-dimensional CW complex. Let g_s and g_t be discrete Morse functions on G, and let V_s and V_t denote their induced discrete gradient vector fields. Suppose that both V_s and V_t are H_0-faithful, meaning that each has exactly one critical vertex in each connected component of G. Let*

$$E = E_s^{\mathrm{crit}} \sqcup E_s^{\mathrm{pair}}, \qquad E = E_t^{\mathrm{crit}} \sqcup E_t^{\mathrm{pair}}$$

be the corresponding spatial and temporal edge decompositions, where $E_\bullet^{\mathrm{pair}}$ denotes the set of edges paired with an incident vertex and $E_\bullet^{\mathrm{crit}}$ denotes the set of unpaired critical edges. Define the four-way partition

$$E_{cc} = E_s^{\mathrm{crit}} \cap E_t^{\mathrm{crit}}, \qquad E_{cp} = E_s^{\mathrm{crit}} \cap E_t^{\mathrm{pair}},$$
$$E_{pc} = E_s^{\mathrm{pair}} \cap E_t^{\mathrm{crit}}, \qquad E_{pp} = E_s^{\mathrm{pair}} \cap E_t^{\mathrm{pair}}.$$

Then

$$|E_{cc}| + |E_{cp}| = |E_{cc}| + |E_{pc}| = \beta_1(G),$$

and hence

$$|E_{cp}| = |E_{pc}|.$$

Proof. Since G is treated as a one-dimensional CW complex, the only cells are nodes and edges. For any discrete Morse function on G, let c_0 denote the number of critical nodes and c_1 the number of critical edges. By the Euler characteristic identity for discrete Morse functions on finite CW complexes,

$$\sum_p (-1)^p c_p = \chi(G).$$

For a one-dimensional complex, this becomes

$$c_0 - c_1 = \chi(G).$$

Since G is a finite graph,

$$\chi(G) = \beta_0(G) - \beta_1(G).$$

Thus

$$c_0 - c_1 = \beta_0(G) - \beta_1(G).$$

Because the induced gradient vector field is H_0-faithful, it has exactly one critical vertex in each connected component of G. Therefore

$$c_0 = \beta_0(G).$$

Substituting into the Euler identity gives

$$\beta_0(G) - c_1 = \beta_0(G) - \beta_1(G),$$

and hence

$$c_1 = \beta_1(G).$$

Therefore, for *any* H_0-faithful discrete gradient vector field on G, the number of critical edges equals $\beta_1(G)$. Applying this conclusion to V_s and V_t separately gives

$$|E_s^{\mathrm{crit}}| = \beta_1(G), \qquad |E_t^{\mathrm{crit}}| = \beta_1(G).$$

By definition of the four-way partition,

$$E_s^{\mathrm{crit}} = E_{cc} \sqcup E_{cp}, \qquad E_t^{\mathrm{crit}} = E_{cc} \sqcup E_{pc}.$$

Therefore

$$|E_{cc}| + |E_{cp}| = \beta_1(G), \qquad |E_{cc}| + |E_{pc}| = \beta_1(G).$$

Hence

$$|E_{cc}| + |E_{cp}| = |E_{cc}| + |E_{pc}| = \beta_1(G).$$

Subtracting $|E_{cc}|$ from both sides yields

$$|E_{cp}| = |E_{pc}|.$$

□

Remark 1. The balance identity is conditional on H_0-faithfulness of both induced gradient vector fields. This condition is not automatic for arbitrary discrete Morse functions, nor is it automatic for the causal temporal Morse function g_t. In particular, events near the beginning of the observation window may have few or no causal-past neighbors, so their temporal scores can be dominated by boundary effects or tie-breaking noise. Consequently, g_t may produce $c_0 > \beta_0(G)$; for a valid discrete Morse vector field, the Euler identity then forces $c_1 > \beta_1(G)$, so the quantitative balance identity above need not hold. Possible mechanisms for enforcing or verifying H_0-faithfulness of g_t, include a warm-up window, a minimum-history threshold, or a post-hoc critical-vertex check.

E.2 Discrete Morse Functions on Graphs

A discrete Morse function on a graph $G = (V, E)$ assigns a real value $g(v)$ to every node and $g(e)$ to every edge such that the resulting gradient field, the set of node-edge pairs (v, e) with v a face of e and $g(v) \geq g(e)$, is acyclic (for further background, refer to Appendix B). Any such gradient field partitions the edge set into two disjoint populations:

$$E = E^{\text{crit}} \sqcup E^{\text{pair}},$$

where a paired edge $e \in E^{\text{pair}}$ is matched with exactly one of its endpoint nodes (forming a gradient pair), and a critical edge $e \in E^{\text{crit}}$ is an edge that remains unpaired.

In practice, we construct the gradient field via a value-monotone greedy procedure following Kannan et al. [24], restricted to graphs. Given vertex values $g(v)$ with distinct values (ties resolved by infinitesimal jitter, and we iterate over edges in any fixed order. For each edge $e = \{u, v\}$ with $g(u) < g(v)$:

- If the higher-valued endpoint v is unmatched, pair v with e and assign the edge value
$$g(e) = \tfrac{1}{2}\big(g(u) + g(v)\big),$$
so that $g(u) < g(e) < g(v)$. Mark v as matched.
- Otherwise, leave e unmatched (so $e \in E^{\text{crit}}$) and assign $g(e) = g(v) + \varepsilon$ for a small positive ε, ensuring e does not induce a further pairing.

Every matched pair (v, e) thus has v as the higher-valued endpoint of e, and Lemma 1 establishes that the resulting partial matching is acyclic. Acyclicity guarantees that paired edges form a forest, so each critical edge closes exactly one fundamental cycle in G relative to this forest. When the gradient field yields a critical node for each connected component (H_0-faithful), the paired edges form a spanning forest and $|E^{\text{crit}}| = \beta_1(G)$.

Lemma 1 (Acyclicity of value-monotone greedy pairings)**.** *Let $G = (V, E)$ be a finite graph, viewed as a one-dimensional CW complex, and let*

$$g : V \mapsto \mathbb{R}$$

be a vertex score function with distinct values. Suppose a partial matching V_g between nodes and incident edges is constructed by the following value-monotone rule: whenever an edge

$$e = \{u, v\}$$

is matched with one of its endpoints, it is matched with the higher-valued endpoint. Equivalently, if

$$(v, e) \in V_g \qquad \textit{and} \qquad e = \{u, v\},$$

then

$$g(u) < g(v).$$

Then V_g is an acyclic discrete gradient vector field on G.

Proof. Since V_g is a partial matching between nodes and incident edges, each cell of G appears in at most one matched pair. It remains to prove that this matching contains no nontrivial closed V_g-path.

On a graph, a nontrivial closed V_g-path would consist of a cyclic sequence

$$v_0, e_0, v_1, e_1, \ldots, v_{r-1}, e_{r-1}, v_r = v_0,$$

where each

$$e_i = \{v_i, v_{i+1}\},$$

and each edge e_i is paired with v_i. Thus following the V_g-path from v_i through the matched edge e_i moves to the opposite endpoint v_{i+1}.

By the value-monotone hypothesis, whenever $e_i = \{v_i, v_{i+1}\}$ is paired with v_i, the opposite endpoint has strictly smaller vertex score:

$$g(v_{i+1}) < g(v_i).$$

Therefore a closed V_g-path would imply

$$g(v_1) < g(v_0), \qquad g(v_2) < g(v_1), \qquad \ldots, \qquad g(v_r) < g(v_{r-1}).$$

Since $v_r = v_0$, this gives the impossible strict cycle of inequalities

$$g(v_0) > g(v_1) > \cdots > g(v_{r-1}) > g(v_0).$$

Hence no nontrivial closed V_g-path exists.

Therefore V_g is an acyclic partial matching, and hence defines a discrete gradient vector field on G. □

Remark 2 (Application to the greedy construction). The greedy procedure described above, following Kannan et al. [24], satisfies the hypothesis of Lemma 1. For each edge $e = \{u, v\}$ with $g(u) < g(v)$, the procedure pairs the higher-valued endpoint v with e (when v is unmatched) and assigns $g(e) = \frac{1}{2}(g(u) + g(v))$; otherwise e is left unmatched. Every matched pair (v, e) therefore has v as the higher-valued endpoint, so the construction is value-monotone. The hypothesis of distinct vertex values is satisfied by the jitter terms $\varepsilon_i^{(s)}$ and $\varepsilon_i^{(t)}$; equivalently, any deterministic tie-breaking suffices. Consequently, the gradient vector fields induced by g_s and g_t—and more generally by any vertex score function used with this procedure—are acyclic.

E.3 Sign-Stability of the Fixed-Graph Morse Skeleton

Proposition 3 (Sign-stability of the fixed-graph Morse skeleton)**.** *Let $S = \{x_1, \ldots, x_n\}$ be an STPP realization with fixed event graph $G(S) = G = (V, E)$, where $V = \{1, \ldots, n\}$. Fix a deterministic edge-processing order $E = \{e_1, \ldots, e_m\}$ and a deterministic tie-breaking convention. Let*

$$g_s(\cdot; S), \qquad g_t(\cdot; S)$$

denote the spatial and temporal vertex score functions used to construct the two discrete Morse pairings on G, and assume both pairings are produced by the value-monotone edge-iteration procedure of Section E.2: when processing an edge $e = \{i, j\}$, the algorithm identifies the higher-valued endpoint and pairs e with that endpoint if it is still unmatched; otherwise e is left unpaired (or assigned a non-pairing edge value according to the fixed convention).

For each edge $e = \{i, j\} \in E$, define the endpoint gaps

$$\Delta_e^{(s)}(S) = g_s(x_i; S) - g_s(x_j; S), \qquad \Delta_e^{(t)}(S) = g_t(x_i; S) - g_t(x_j; S),$$

and assume S is a no-tie state:

$$\Delta_e^{(s)}(S) \neq 0, \quad \Delta_e^{(t)}(S) \neq 0 \quad \textit{for all } e \in E.$$

The fixed-graph comparison margin is

$$m_{\mathrm{comp}}(S) := \min_{e \in E} \left\{ |\Delta_e^{(s)}(S)|,\ |\Delta_e^{(t)}(S)| \right\}.$$

Let $S' = \{x'_1, \ldots, x'_n\}$ be another realization on the same vertex set V, satisfying the structural hypotheses

$$G(S') = G(S) = G, \qquad t_i < t_j \iff t'_i < t'_j \quad \textit{for all } i, j,$$

and the metric hypotheses

$$|\Delta_e^{(s)}(S') - \Delta_e^{(s)}(S)| < m_{\mathrm{comp}}(S), \qquad |\Delta_e^{(t)}(S') - \Delta_e^{(t)}(S)| < m_{\mathrm{comp}}(S)$$

for every $e \in E$. Then

$$\mathrm{Skel}(S') = \mathrm{Skel}(S).$$

Proof. Fix $\bullet \in \{s, t\}$ and write $\Delta_e^{(\bullet)}(S) = g_\bullet(x_i; S) - g_\bullet(x_j; S)$ for $e = \{i, j\} \in E$. The metric hypothesis gives

$$|\Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S)| < m_{\mathrm{comp}}(S) \le |\Delta_e^{(\bullet)}(S)|,$$

so $\Delta_e^{(\bullet)}(S')$ lies in an open interval around $\Delta_e^{(\bullet)}(S)$ that excludes zero. Hence

$$\operatorname{sign} \Delta_e^{(\bullet)}(S') = \operatorname{sign} \Delta_e^{(\bullet)}(S) \qquad \text{for every } e \in E,$$

and therefore the same endpoint of e is higher-valued in both runs.

The pairings agree. Let $M_k^{(\bullet)}(S) \subseteq V \times E$ denote the partial matching obtained from $g_\bullet(\cdot; S)$ after processing edges $e_1, \ldots, e_k$, and define $M_k^{(\bullet)}(S')$ analogously. We show by induction on k that $M_k^{(\bullet)}(S) = M_k^{(\bullet)}(S')$ for $k = 0, 1, \ldots, m$.

The case $k = 0$ is immediate, since $M_0^{(\bullet)}(S) = \varnothing = M_0^{(\bullet)}(S')$. For the inductive step, suppose $M_{k-1}^{(\bullet)}(S) = M_{k-1}^{(\bullet)}(S')$, and let $e_k = \{i, j\}$. The endpoint-sign preservation established above identifies a common higher-valued endpoint $h_k^{(\bullet)} \in \{i, j\}$ in both runs, and the induction hypothesis ensures that $h_k^{(\bullet)}$ has the same matched/unmatched status in $M_{k-1}^{(\bullet)}(S)$ and $M_{k-1}^{(\bullet)}(S')$. The value-monotone rule therefore takes the same action at step k in both runs:

$$M_k^{(\bullet)}(S) = \begin{cases} M_{k-1}^{(\bullet)}(S) \cup \{(h_k^{(\bullet)}, e_k)\}, & \text{if } h_k^{(\bullet)} \text{ is unmatched,} \\ M_{k-1}^{(\bullet)}(S), & \text{if } h_k^{(\bullet)} \text{ is matched,} \end{cases}$$

and the same identity holds for S'. Hence $M_k^{(\bullet)}(S) = M_k^{(\bullet)}(S')$, completing the induction.

The four-way partition agrees. Applying the conclusion to $\bullet = s$ and $\bullet = t$ gives $M_m^{(s)}(S) = M_m^{(s)}(S')$ and $M_m^{(t)}(S) = M_m^{(t)}(S')$, so

$$E_s^{\mathrm{pair}}(S) = E_s^{\mathrm{pair}}(S'), \qquad E_t^{\mathrm{pair}}(S) = E_t^{\mathrm{pair}}(S').$$

Since $E_\bullet^{\mathrm{crit}} = E \setminus E_\bullet^{\mathrm{pair}}$, the critical edge sets agree as well. Each block of the four-way partition is an intersection of a spatial set and a temporal set, both of which are invariant under $S \to S'$; hence

$$E_{cc}(S) = E_{cc}(S'), \quad E_{cp}(S) = E_{cp}(S'), \quad E_{pc}(S) = E_{pc}(S'), \quad E_{pp}(S) = E_{pp}(S').$$

Therefore $\mathrm{Skel}(S') = \mathrm{Skel}(S)$. The numerical vertex and edge values assigned by the procedure may differ between S and S', since they depend on the actual scores rather than only their signs; what is preserved is the four-way critical/paired edge assignment. □

Remark 3 (Static comparison set). Under the value-monotone edge-iteration procedure, the fixed-graph extractor is determined by the $2|E|$ endpoint-ordering signs $\operatorname{sign} \Delta_e^{(s)}$ and $\operatorname{sign} \Delta_e^{(t)}$ for $e \in E$. The matched/unmatched status of a vertex at step k is not an additional comparison: it is a deterministic function of the preceding signs and the fixed edge order, as the induction in the proof shows. The comparison set is therefore static, fixed by G alone.

Remark 4 (Scope and limitations). The structural hypotheses $G(S') = G(S)$ and $t_i < t_j \iff t'_i < t'_j$ are assumptions of the proposition; they are not consequences of the comparison margin. A full raw-data perturbation theorem would require additional structural margins—a k-nearest-neighbor margin to ensure graph preservation, and a temporal-order margin to ensure preservation of causal relations—under which an arbitrarily small perturbation cannot change a nearest-neighbor relation or swap nearly tied timestamps. The development of such margins, and the resulting joint stability theorem, is left to future work.

E.4 Lipschitz Bound for Kernel Morse Scores and Fixed-Graph Stability Radius

Proposition 4 (Fixed-graph Lipschitz bound for kernel Morse scores)**.** *Let $S = \{x_1, \dots, x_n\}$, with $x_i = (t_i, s_i)$, be an STPP realization, and let*

$$G(S) = G = (V, E), \qquad V = \{1, \dots, n\}.$$

Let $S' = \{x'_1, \dots, x'_n\}$, with $x'_i = (t'_i, s'_i)$, be another realization on the same vertex set. Assume the structural hypotheses

$$G(S') = G(S) = G$$

and

$$t_i < t_j \iff t'_i < t'_j \qquad \text{for all } i, j.$$

Assume the normalized perturbation bound

$$\max_i \left[\left(\frac{t_i - t'_i}{\sigma_t} \right)^2 + \left(\frac{\|s_i - s'_i\|}{\sigma_s} \right)^2 \right]^{1/2} \le \rho.$$

Let $K : \mathbb{R}_{\ge 0} \to \mathbb{R}$ be a Lipschitz kernel with

$$\|K'\|_\infty \le L_K.$$

Consider spatial and temporal Morse scores

$$g_s(x_i; S) = -\sum_{j \in N_G(i)} w_{ij}^{(s)} K\left(\frac{\|s_i - s_j\|}{\sigma_s} \right) + \varepsilon_i^{(s)}$$

and

$$g_t(x_i; S) = -\sum_{j \in N_G^-(i)} w_{ij}^{(t)} K\left(\frac{|t_i - t_j|}{\sigma_t} \right) + \varepsilon_i^{(t)},$$

where

$$N_G^-(i) := \{ j \in N_G(i) : t_j < t_i \}.$$

Assume the weights $w_{ij}^{(s)}, w_{ij}^{(t)}$ and jitter terms $\varepsilon_i^{(s)}, \varepsilon_i^{(t)}$ are fixed under the perturbation. Define

$$A_s := \max_i \sum_{j \in N_G(i)} |w_{ij}^{(s)}|, \qquad A_t := \max_i \sum_{j \in N_G^-(i)} |w_{ij}^{(t)}|,$$

and

$$A_{\max} := \max\{A_s, A_t\}.$$

For each edge $e = \{i, j\} \in E$, define

$$\Delta_e^{(s)}(S) = g_s(x_i; S) - g_s(x_j; S), \qquad \Delta_e^{(t)}(S) = g_t(x_i; S) - g_t(x_j; S).$$

Then, for every $e \in E$,

$$\left| \Delta_e^{(s)}(S') - \Delta_e^{(s)}(S) \right| \le 4 L_K A_s \rho,$$

and

$$\left| \Delta_e^{(t)}(S') - \Delta_e^{(t)}(S) \right| \le 4 L_K A_t \rho.$$

Consequently,

$$\max_{\bullet \in \{s,t\}} \max_{e \in E} \left| \Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S) \right| \le 4 L_K A_{\max} \rho.$$

Proof. The normalized perturbation assumption implies, for every i,

$$\|s_i - s'_i\| \le \sigma_s \rho, \qquad |t_i - t'_i| \le \sigma_t \rho.$$

Hence, for every pair i, j,

$$\begin{aligned}\left|\frac{\|s_i' - s_j'\|}{\sigma_s} - \frac{\|s_i - s_j\|}{\sigma_s}\right| &\le \frac{\|(s_i' - s_j') - (s_i - s_j)\|}{\sigma_s} \\ &\le \frac{\|s_i' - s_i\| + \|s_j' - s_j\|}{\sigma_s} \\ &\le 2\rho.\end{aligned}$$

Similarly,

$$\begin{aligned}\left|\frac{|t_i' - t_j'|}{\sigma_t} - \frac{|t_i - t_j|}{\sigma_t}\right| &\le \frac{|(t_i' - t_j') - (t_i - t_j)|}{\sigma_t} \\ &\le \frac{|t_i' - t_i| + |t_j' - t_j|}{\sigma_t} \\ &\le 2\rho.\end{aligned}$$

Since K is L_K-Lipschitz,

$$\left|K\left(\frac{\|s_i' - s_j'\|}{\sigma_s}\right) - K\left(\frac{\|s_i - s_j\|}{\sigma_s}\right)\right| \le 2L_K\rho,$$

and

$$\left|K\left(\frac{|t_i' - t_j'|}{\sigma_t}\right) - K\left(\frac{|t_i - t_j|}{\sigma_t}\right)\right| \le 2L_K\rho.$$

For the spatial score,

$$\begin{aligned}|g_s(x_i; S') - g_s(x_i; S)| &\le \sum_{j \in N_G(i)} |w_{ij}^{(s)}| \left|K\left(\frac{\|s_i' - s_j'\|}{\sigma_s}\right) - K\left(\frac{\|s_i - s_j\|}{\sigma_s}\right)\right| \\ &\le 2L_K\rho \sum_{j \in N_G(i)} |w_{ij}^{(s)}| \\ &\le 2L_K A_s \rho.\end{aligned}$$

Therefore, for $e = \{i, j\}$,

$$\begin{aligned}\left|\Delta_e^{(s)}(S') - \Delta_e^{(s)}(S)\right| &= \left|\big(g_s(x_i; S') - g_s(x_j; S')\big) - \big(g_s(x_i; S) - g_s(x_j; S)\big)\right| \\ &\le |g_s(x_i; S') - g_s(x_i; S)| + |g_s(x_j; S') - g_s(x_j; S)| \\ &\le 4L_K A_s \rho.\end{aligned}$$

For the temporal score, the causal-order hypothesis gives

$$N_G^-(i; S') = N_G^-(i; S) \qquad \text{for all } i.$$

Thus the same argument gives

$$\begin{aligned}|g_t(x_i; S') - g_t(x_i; S)| &\le \sum_{j \in N_G^-(i)} |w_{ij}^{(t)}| \left|K\left(\frac{|t_i' - t_j'|}{\sigma_t}\right) - K\left(\frac{|t_i - t_j|}{\sigma_t}\right)\right| \\ &\le 2L_K\rho \sum_{j \in N_G^-(i)} |w_{ij}^{(t)}| \\ &\le 2L_K A_t \rho.\end{aligned}$$

Therefore,

$$\left|\Delta_e^{(t)}(S') - \Delta_e^{(t)}(S)\right| \le 4L_K A_t \rho.$$

Taking the maximum over $e \in E$ and $\bullet \in \{s, t\}$ gives

$$\max_{\bullet \in \{s,t\}} \max_{e \in E} \left|\Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S)\right| \le 4L_K A_{\max} \rho.$$

□

Corollary 1 (Fixed-graph perturbation stability)**.** *Assume the fixed event graph and causal order are preserved:*

$$G(S') = G(S) = G, \qquad t_i < t_j \iff t'_i < t'_j \qquad \textit{for all } i, j.$$

Assume the spatial and temporal pairings are constructed by the value-monotone edge-iteration procedure of Section E.2, and that S is a no-tie state:

$$\Delta_e^{(s)}(S) \neq 0, \qquad \Delta_e^{(t)}(S) \neq 0 \qquad \textit{for all } e \in E.$$

Assume further that g_s and g_t are kernel Morse scores of the form in Proposition 4, with Lipschitz kernel constant L_K, fixed weights and jitter terms, and aggregation constant

$$A_{\max} := \max\{A_s, A_t\}.$$

Let

$$m_{\mathrm{comp}}(S) := \min_{e \in E} \left\{ |\Delta_e^{(s)}(S)|,\ |\Delta_e^{(t)}(S)| \right\}.$$

If the normalized perturbation magnitude satisfies

$$\rho < \frac{m_{\mathrm{comp}}(S)}{4 L_K A_{\max}},$$

then

$$\mathrm{Skel}(S') = \mathrm{Skel}(S).$$

Proof. By Proposition 4,

$$\max_{\bullet \in \{s,t\}} \max_{e \in E} \left| \Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S) \right| \leq 4 L_K A_{\max} \rho.$$

The radius condition implies

$$4 L_K A_{\max} \rho < m_{\mathrm{comp}}(S).$$

Hence, for every $e \in E$ and $\bullet \in \{s, t\}$,

$$\left| \Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S) \right| < m_{\mathrm{comp}}(S).$$

Thus the comparison-gap hypotheses of Proposition 3 hold. Since the structural, procedural, and no-tie hypotheses of Proposition 3 are also assumed above, we obtain

$$\mathrm{Skel}(S') = \mathrm{Skel}(S).$$

□

Remark 5 (Normalized versus unnormalized perturbations). The perturbation parameter ρ in Proposition 4 is normalized by the bandwidths σ_t and σ_s. This is why no additional factors of $1/\sigma_t$ or $1/\sigma_s$ appear in the Lipschitz constant.

If instead one measures perturbations in unnormalized units by

$$\rho_{\mathrm{raw}} := \max_i \left[|t_i - t'_i|^2 + \|s_i - s'_i\|^2 \right]^{1/2},$$

then

$$\rho \leq \frac{\rho_{\mathrm{raw}}}{\min\{\sigma_t, \sigma_s\}}.$$

Consequently,

$$\max_{\bullet \in \{s,t\}} \max_{e \in E} \left| \Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S) \right| \leq \frac{4 L_K A_{\max}}{\min\{\sigma_t, \sigma_s\}} \rho_{\mathrm{raw}}.$$

Equivalently, the fixed-graph stability radius in raw units is at least

$$\rho_{\mathrm{raw}} < \frac{m_{\mathrm{comp}}(S) \min\{\sigma_t, \sigma_s\}}{4 L_K A_{\max}}.$$

Remark 6 (Aggregation constants). For unweighted neighbor-sum scores ($w_{ij}^{(s)} = w_{ij}^{(t)} = 1$),

$$A_s = \Delta_G, \qquad A_t = \max_i |N_G^-(i)| \le \Delta_G, \qquad \Delta_G := \max_{i \in V} \deg_G(i),$$

hence $A_{\max} = \Delta_G$. For normalized neighbor averages, where

$$\sum_{j \in N_G(i)} |w_{ij}^{(s)}| \le 1, \qquad \sum_{j \in N_G^-(i)} |w_{ij}^{(t)}| \le 1,$$

one may take $A_{\max} \le 1$; equality holds in the usual uniform-average case. For a symmetrized k-nearest-neighbor graph, Δ_G should be treated as the actual maximum degree of the symmetrized graph; it is not generally equal to k.

Remark 7 (Gaussian and Matérn kernels). Write

$$L_{\text{comp}} := 4 L_K A_{\max}.$$

For the Gaussian kernel $K(r) = \exp(-r^2/2)$,

$$\|K'\|_\infty = e^{-1/2},$$

so

$$L_{\text{comp}} = 4 e^{-1/2} A_{\max}.$$

For unweighted neighbor-sum Gaussian Morse scores, $A_{\max} = \Delta_G$ by the previous remark, giving

$$L_{\text{comp}} = 4 e^{-1/2} \Delta_G.$$

More generally, Proposition 4 applies to any Lipschitz kernel, including the standard Matérn choices $\nu = 1/2, 3/2, 5/2$. Discontinuous kernels such as the Tophat kernel are excluded from this Lipschitz argument.

The decision-gap Lipschitz constant of Definition 2 depends on ν through the maximum kernel slope $\|K'_\nu\|_\infty$. Under the normalized perturbation model of Definition 4,

$$L_s(\nu) \;\le\; 4 \|K'_\nu\|_\infty A_s, \qquad L_t(\nu) \;\le\; 4 \|K'_\nu\|_\infty A_t, \tag{15}$$

where the aggregation constants are $A_s := \max_i \sum_{j \in \mathcal{N}_G(i)} |w_{ij}^{(s)}|$ and $A_t := \max_i \sum_{j \in \mathcal{N}_G^-(i)} |w_{ij}^{(t)}|$, with $\mathcal{N}_G^-(i) := \{j \in \mathcal{N}_G(i) : t_j < t_i\}$ the causal-past neighborhood. For unweighted neighbor-sum scores ($w_{ij}^{(s)} = w_{ij}^{(t)} = 1$), $A_s = \Delta_G$ and $A_t \le \Delta_G$, where $\Delta_G := \max_{i \in V} \deg_G(i)$ is the maximum degree of the symmetrized kNN graph. We emphasize that Δ_G is not generally equal to k: although each vertex has exactly k outgoing kNN edges, symmetrization can produce nodes of substantially higher degree in dense regions. The factor of 4 arises as 2×2: one factor of 2 from the bandwidth-normalized triangle inequality on each pair of perturbed events, and a second factor of 2 from the gap $\Delta_e = g(x_i) - g(x_j)$ involving two endpoints (Appendix E). The slope $\|K'_\nu\|_\infty$ is monotonically non-increasing in ν on $\nu \ge 1/2$:

$$\|K'_{1/2}\|_\infty = 1, \quad \|K'_{3/2}\|_\infty = \sqrt{3}\, e^{-1} \approx 0.638, \quad \|K'_\infty\|_\infty = e^{-1/2} \approx 0.607. \tag{16}$$

Hence higher Matérn smoothness yields a smaller Lipschitz constant and a larger stability radius. We define the unified decision-gap Lipschitz constant

$$L(\nu) \;:=\; 4 \|K'_\nu\|_\infty A_{\max}, \qquad A_{\max} \;:=\; \max\{A_s, A_t\}.$$

E.5 Raw-Data Perturbation Stability

Corollary 1 certifies skeleton preservation under perturbations that already preserve the event graph and causal order. We now lift that result to a raw-data perturbation theorem: under additional margin conditions on the k-nearest-neighbor structure and on temporal ordering, sufficiently small perturbations of the raw event coordinates preserve the skeleton.

For arbitrary space-time points $x = (t_x, s_x)$ and $y = (t_y, s_y)$, define the bandwidth-normalized distance

$$d_\sigma(x, y) := \left[\frac{(t_x - t_y)^2}{\sigma_t^2} + \frac{\|s_x - s_y\|^2}{\sigma_s^2} \right]^{1/2},$$

which is a metric on $\mathbb{R} \times \mathbb{R}^d$ for positive bandwidths $\sigma_s, \sigma_t > 0$. For an STPP realization $S = \{x_1, \ldots, x_n\}$, write $d_S(i,j) := d_\sigma(x_i, x_j)$; for a second realization $S' = \{x'_1, \ldots, x'_n\}$, write $d_{S'}(i,j) := d_\sigma(x'_i, x'_j)$. Let $G(S)$ be the symmetrized directed k-nearest-neighbor graph under d_S, and similarly $G(S')$ under $d_{S'}$. For each i, let

$$d_i^{(1)}(S) \le d_i^{(2)}(S) \le \cdots \le d_i^{(n-1)}(S)$$

denote the ordered distances from x_i to all other events in S.

Definition 25 (Structural margins)**.** Assume $1 \le k \le n-2$. The k-nearest-neighbor margin and temporal-order margin of S are

$$\eta_{\mathrm{kNN}}(S) := \min_i \left(d_i^{(k+1)}(S) - d_i^{(k)}(S) \right), \qquad \eta_{\mathrm{ord}}(S) := \min_{i \ne j} \frac{|t_i - t_j|}{\sigma_t}.$$

Definition 26 (Raw-data stability radius)**.** The raw-data stability radius of S is

$$r_{\mathrm{raw}}(S) := \min \left\{ \frac{\eta_{\mathrm{kNN}}(S)}{4}, \ \frac{\eta_{\mathrm{ord}}(S)}{2}, \ \frac{m_{\mathrm{comp}}(S)}{4 L_K A_{\max}} \right\},$$

where $m_{\mathrm{comp}}(S)$ is the fixed-graph Morse comparison margin from Proposition 3, L_K is the kernel Lipschitz constant, and $A_{\max}$ is the aggregation constant from Proposition 4.

Theorem 2 (Raw-data perturbation stability of the Morse skeleton)**.** *Let $S = \{x_1, \ldots, x_n\}$, with $x_i = (t_i, s_i)$, be an STPP realization with event graph $G(S)$ constructed as above. Assume the following hypotheses.*

Structural margins: *the k-nearest-neighbor and temporal-order margins are positive,*

$$\eta_{\mathrm{kNN}}(S) > 0, \qquad \eta_{\mathrm{ord}}(S) > 0.$$

Morse no-tie: *S is a no-tie state for the Morse comparison gaps,*

$$\Delta_e^{(s)}(S) \ne 0, \quad \Delta_e^{(t)}(S) \ne 0 \quad \textit{for all } e \in E(G(S)),$$

so that the comparison margin $m_{\mathrm{comp}}(S) > 0$.

Score form: *the spatial and temporal Morse scores g_s, g_t are kernel Morse scores of the form in Proposition 4, with Lipschitz kernel constant L_K and aggregation constant $A_{\max} = \max\{A_s, A_t\}$. The bandwidths σ_s, σ_t and kernel weights $w_{ij}^{(s)}, w_{ij}^{(t)}$ are fixed, and the jitter terms $\varepsilon_i^{(s)}, \varepsilon_i^{(t)}$ are sampled once and held fixed under the perturbation. The pairings are constructed by the value-monotone edge-iteration procedure of Section E.2.*

Perturbation: *$S' = \{x'_1, \ldots, x'_n\}$, with $x'_i = (t'_i, s'_i)$, is a pointwise perturbation of S (the vertex indexing is fixed), satisfying*

$$\max_i d_\sigma(x_i, x'_i) \le \rho.$$

If

$$\rho < r_{\mathrm{raw}}(S),$$

then the Morse skeleton produced by the value-monotone construction is unchanged:

$$\mathrm{Skel}(S') = \mathrm{Skel}(S).$$

Proof. The proof proceeds in three steps: graph preservation, causal-order preservation, and Morse-comparison-sign preservation. Once these are established, Proposition 3 delivers the conclusion.

Step 1: $G(S') = G(S)$. For any pair i, j, the triangle inequality for the metric d_σ gives

$$|d_{S'}(i,j) - d_S(i,j)| = |d_\sigma(x'_i, x'_j) - d_\sigma(x_i, x_j)| \le d_\sigma(x_i, x'_i) + d_\sigma(x_j, x'_j) \le 2\rho.$$

Fix i. Let $j \in \mathrm{kNN}_k(i; S)$ be a directed k-nearest neighbor of i in S, and let $\ell \notin \mathrm{kNN}_k(i; S)$. Then

$$d_S(i,j) \le d_i^{(k)}(S), \qquad d_S(i,\ell) \ge d_i^{(k+1)}(S),$$

so

$$d_S(i,\ell) - d_S(i,j) \ge \eta_{\mathrm{kNN}}(S).$$

Under perturbation,

$$d_{S'}(i,\ell) - d_{S'}(i,j) \geq \big(d_S(i,\ell) - 2\rho\big) - \big(d_S(i,j) + 2\rho\big) = d_S(i,\ell) - d_S(i,j) - 4\rho \geq \eta_{\mathrm{kNN}}(S) - 4\rho.$$

Since $\rho < \eta_{\mathrm{kNN}}(S)/4$, we have $d_{S'}(i,\ell) > d_{S'}(i,j)$. Hence every directed k-nearest neighbor of i in S remains strictly closer than every non-neighbor, and the directed sets satisfy $\mathrm{kNN}_k(i;S') = \mathrm{kNN}_k(i;S)$ for every i. Since the symmetrized graph is defined by

$$\{i,j\} \in E(G(S)) \iff j \in \mathrm{kNN}_k(i;S) \text{ or } i \in \mathrm{kNN}_k(j;S),$$

preservation of every directed k-NN set implies preservation of the symmetrized graph: $G(S') = G(S)$.

Step 2: causal order is preserved. For every pair i, j,

$$\frac{|(t'_i - t'_j) - (t_i - t_j)|}{\sigma_t} \leq \frac{|t'_i - t_i|}{\sigma_t} + \frac{|t'_j - t_j|}{\sigma_t} \leq 2\rho.$$

Since $\rho < \eta_{\mathrm{ord}}(S)/2$ and $\eta_{\mathrm{ord}}(S) \leq |t_i - t_j|/\sigma_t$ for every $i \neq j$, no temporal difference can change sign:

$$t_i < t_j \iff t'_i < t'_j \qquad \text{for all } i, j.$$

In particular, $N_G^-(i;S') = N_G^-(i;S)$ for every i, so the aggregation constant A_t and the temporal Morse score's neighbor index sets are unchanged.

Step 3: Morse comparison signs are preserved. Steps 1 and 2 establish the structural hypotheses required by Proposition 4. Applying the proposition under these now-verified hypotheses gives

$$\max_{\bullet \in \{s,t\}} \max_{e \in E(G(S))} \left| \Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S) \right| \leq 4 L_K A_{\max} \rho.$$

Since $\rho < m_{\mathrm{comp}}(S)/(4L_K A_{\max})$, this gives

$$\left| \Delta_e^{(\bullet)}(S') - \Delta_e^{(\bullet)}(S) \right| < m_{\mathrm{comp}}(S)$$

for every $e \in E(G(S))$ and $\bullet \in \{s,t\}$.

Conclusion. Steps 1, 2, and 3 verify the structural and metric hypotheses of Proposition 3. The procedural and no-tie hypotheses are assumed in the theorem statement. Therefore Proposition 3 applies, giving $\mathrm{Skel}(S') = \mathrm{Skel}(S)$. □

Remark 8 (Vacuity of the radius). The raw-data stability radius $r_{\mathrm{raw}}(S)$ is positive only when the structural and comparison margins are all positive. The condition $\eta_{\mathrm{kNN}}(S) > 0$ requires that no event has its k-th and $(k+1)$-th nearest neighbors at exactly equal normalized distance, and $\eta_{\mathrm{ord}}(S) > 0$ requires distinct timestamps. These conditions hold generically but can fail in datasets with discretized timestamps, exactly coincident events, or symmetric configurations. In tied cases, one may impose deterministic tie-breaking or add an infinitesimal preprocessing jitter to select a nearby no-tie realization; the stability certificate then applies around that selected realization, not necessarily around the original tied state. The Morse no-tie condition $m_{\mathrm{comp}}(S) > 0$ is generically satisfied by the jitter terms in both our spatial and temporal Morse functions.

Remark 9 (Structural assumptions held fixed). The theorem assumes a pointwise perturbation $x_i \mapsto x'_i$ under fixed indexing: S' is a coordinate perturbation of S, not a reordering, resampling, or change in event count. The bandwidths (σ_s, σ_t), kernel weights, and jitter terms are also held fixed; the result does not address robustness to bandwidth selection, weight rescaling, or jitter resampling. These are separate questions and are not addressed by the present theorem.

F Visualizations

General prompt

You are a geospatial intelligence expert specializing in spatial pattern analysis and environmental event modeling. Your task is to analyze map-view imagery containing geo-referenced event points and extract structured, semantically meaningful spatial patterns.

Data description

The input consists of a map-view image representing wildfire event occurrences distributed across Southern California. Each point corresponds to an event localized by latitude and longitude. The map overlays these event points on a geographic map, where terrain and regional boundaries are implicitly encoded through the background map structure rather than explicit annotations. The goal is to uncover interpretable spatial organization without access to explicit metadata.

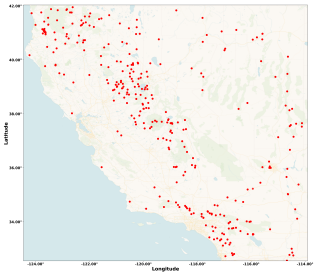


Figure 2: An example prompt for an image of the wildfire dataset.

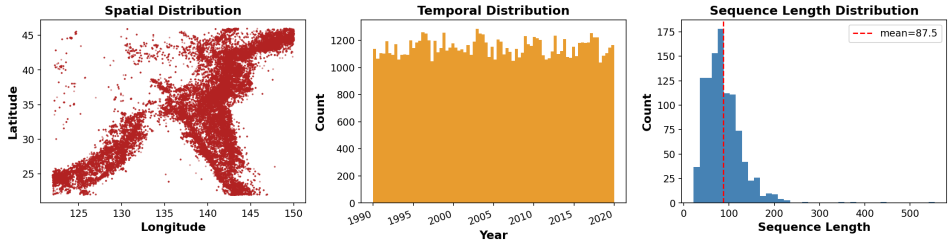


Figure 3: Visualization of Japan earthquake.

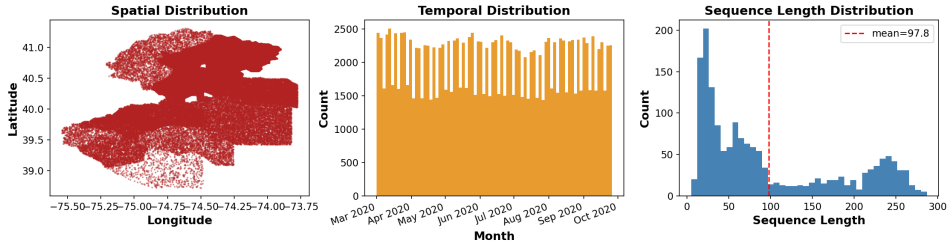


Figure 4: Visualization of COVID-19.

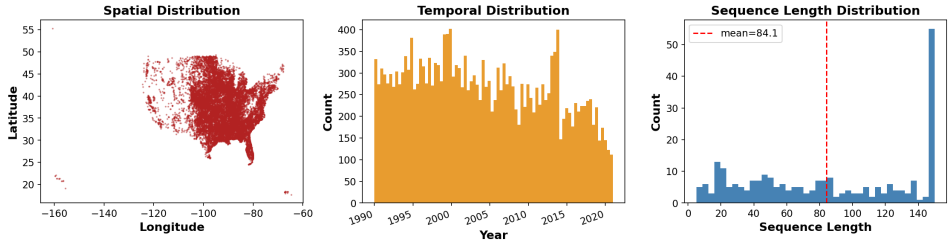


Figure 5: Visualization of US tornado.

G Additional Experimental Results

We search the optimal value of η_1, which controls the multimodal contrastive loss, over the range $[0.01, 0.1]$. This range is chosen to ensure that the contrastive objective provides sufficient cross-modal alignment while avoiding domination over the main forecasting loss. Empirically, we found $\eta_1 = 0.05$ provides a stable balance between representation consistency and predictive flexibility.

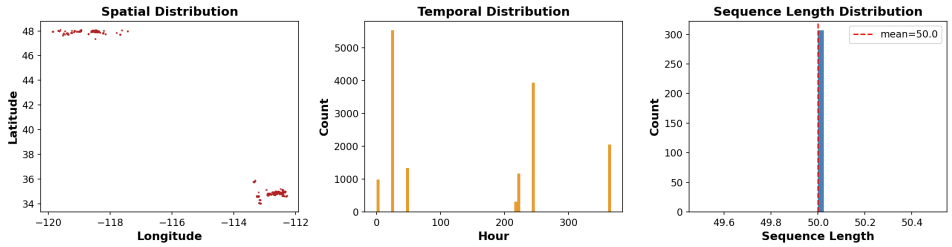


Figure 6: Visualization of eMAS data.

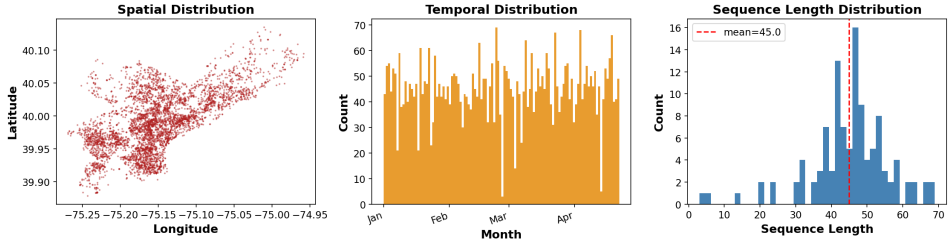


Figure 7: Visualization of crime data.

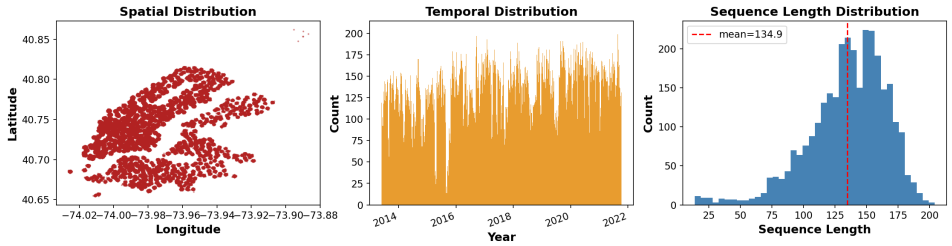


Figure 8: Visualization of Citibike data.

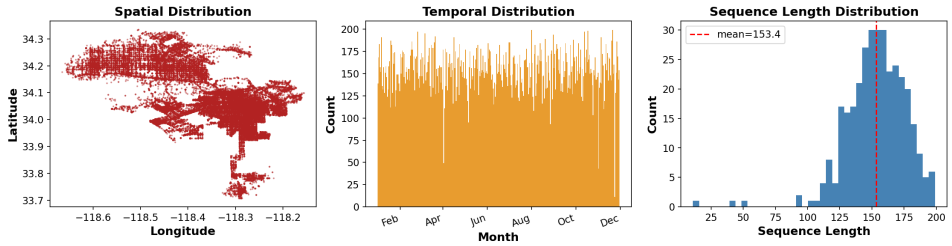


Figure 9: Visualization of traffic collision data.

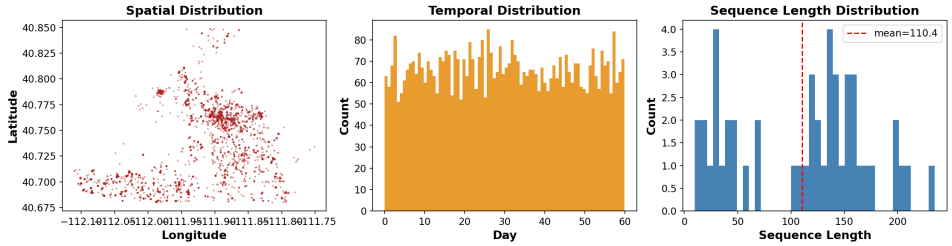


Figure 10: Visualization of twitter data.

The sensitivity analysis with respect to η_2, η_3 reveals a clear trade-off in balancing the Morse-based spatial and temporal loss terms. We observe that moderate values consistently achieve the best or near-best performance across both datasets, indicating that a mild topological constraint is sufficient to guide the model toward meaningful structural representations. When η_2 and η_3 become large, performance degrades noticeably (particularly in the temporal domain on Twitter) suggesting that overly strong Morse-based regularization may over-constrain the learning dynamics and hinder the model's ability to adapt to data-driven patterns. This behavior highlights that while the Morse-based spatial and temporal losses play a critical role in capturing higher-order topological structures, their contribution is most effective when used as a gentle inductive bias rather than a dominant optimization objective.

Table 6: Hyperparameter Sensitivity Analysis.

	Crime		Twitter	
η_2, η_3	Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓
0.0005	0.0438±0.0008	**0.0305±0.0005**	0.0706±0.0116	**0.8472±0.0379**
0.001	**0.0435±0.0010**	0.0315±0.0004	**0.0690±0.0079**	0.8683±0.0337
0.005	0.0497±0.0008	0.0323±0.0008	0.0818±0.0023	0.9551±0.0849
0.01	0.0597±0.0032	0.0346±0.0002	0.0855±0.0014	1.0779±0.1067
0.05	0.0685±0.0004	0.0511±0.0063	0.0856±0.0015	2.5867±0.6310
0.1	0.0700±0.0018	0.0727±0.0053	0.0848±0.0019	3.8050±0.7434

We have conducted sensitivity analyses covering the major hyperparameters including the loss term weight η_1, diffusion timesteps, and noise schedule. As shown in Tables 7 and 8, the performance varies only modestly with different values of η_1. Similarly, the diffusion model exhibits stable behavior under different diffusion timesteps and both cosine and linear noise schedules.

Table 9 evaluates robustness under distributional perturbations by injecting 5% and 10% Gaussian noise (with $\mu = 0$ and $\sigma = 1$) into both training and validation sets. Across all settings, MCM-DM consistently outperforms TEN-DM on both spatial and temporal metrics. Under 5% noise, MCM-DM achieves substantial relative improvements on the US Tornado dataset, reducing spatial error by 20.26% and temporal error by 5.27% compared to TEN-DM. Similar gains are observed on the Crime dataset, where MCM-DM improves spatial and temporal errors by 43.44% and 32.02%, respectively. As noise intensity increases to 10%, both methods experience performance degradation; on US Tornado, MCM-DM maintains a clear advantage over TEN-DM, achieving 17.93% and 11.96% relative improvements in spatial and temporal dimensions, respectively. This consistent gap indicates that MCM-DM is less sensitive to input perturbations and maintains more stable predictions under noisy conditions.

H Datasets, Baselines, and Implementation Details

Datasets: JPN Earthquake: Earthquake events with a magnitude of at least 2.5 in Japan from 1990 to 2020[1]; **COVID-19:** COVID-19 dataset is collected from publicly released COVID-19 cases in New Jersey state from March 2020 to July 2020[2]; **US Tornado:** Tornado occurrence records across the contiguous United States from 1950 to 2021[3]; **eMAS:** Wildfire occurrence records from the Enhanced MODIS Airborne Simulator (eMAS) which are across the western United States over approximately one year; **Crime:** Crime incidents including robbery, burglary and weapon violations reported by the Philadelphia Police Department in Philadelphia in 2026[4]; **Citibike:** Bike-sharing trip records from the New York City Citi Bike system in the Manhattan area from 2013 to 2022 [54]; **Traffic Collision:** Traffic collision records in Los Angeles from 2010 to 2019[5]; **Twitter:** Geotagged Twitter posts collected in the Salt Lake City area [27].

Japan Earthquake: The dataset is partitioned into nonoverlapping splits containing 950 training sequences, 50 validation sequences, and 50 test sequences (train/validation/test: 90%/5%/5%).

[1] https://earthquake.usgs.gov/earthquakes/search/

[2] https://github.com/nytimes/covid-19-data

[3] https://www.kaggle.com/datasets/danbraswell/us-tornado-dataset-1950-2021

[4] https://opendataphilly.org/

[5] https://data.lacity.org/Public-Safety/Traffic-Collision-Data-from-2010-to-Present/d5tf-ez2w/about_data

Table 7: Hyperparameter sensitivity analysis on the Crime.

Hyperparameter	Value	Spatial ↓	Temporal ↓
η_1	0.01	0.0442±0.0004	0.0199±0.0004
	0.05	**0.0434±0.0011**	0.0201±0.0007
	0.1	0.0442±0.0008	0.0201±0.0010
	0.5	0.0454±0.0013	**0.0192±0.0005**
Diffusion timesteps	250	**0.0431±0.0003**	**0.0189±0.0002**
	500	0.0434±0.0011	0.0201±0.0007
	1000	0.0449±0.0004	0.0192±0.0003
Noise schedule	cosine	**0.0434±0.0011**	0.0201±0.0007
	linear	0.0436±0.0003	**0.0195±0.0006**

Table 8: Hyperparameter sensitivity analysis on the Twitter.

Hyperparameter	Value	Spatial ↓	Temporal ↓
η_1	0.01	**0.0442±0.0008**	**0.4020±0.0107**
	0.05	0.0449±0.0015	0.4139±0.0274
	0.1	0.0454±0.0014	0.4066±0.0075
	0.5	0.0468±0.0022	0.4047±0.0034
Diffusion timesteps	250	0.0451±0.0005	0.4331±0.0116
	500	**0.0449±0.0015**	**0.4139±0.0274**
	1000	0.0460±0.0007	0.4232±0.0195
Noise schedule	cosine	0.0449±0.0015	0.4139±0.0274
	linear	**0.0447±0.0006**	**0.3988±0.0210**

Table 9: Robustness Analysis.

Ratio (%)	Method	Crime		US Tornado	
		Spatial ↓	Temporal ↓	Spatial ↓	Temporal ↓
5%	TEN-DM	0.1261±0.0178	0.0641±0.0070	12.6632±0.1810	1.1169±0.0397
	MCM-DM (Ours)	**0.0714±0.0023**	**0.0436±0.0045**	**10.0979±0.1337**	**1.0581±0.0213**
10%	TEN-DM	0.1513±0.0062	0.0898±0.0039	13.5589±0.2498	1.1207±0.0341
	MCM-DM (Ours)	**0.0756±0.0014**	**0.0569±0.0165**	**11.1273±0.1219**	**1.0628±0.0504**

Sequence lengths range from 22 to 554. **COVID-19:** The dataset is partitioned into 1,450 training sequences, 100 validation sequences, and 100 test sequences (train/validation/test: 88%/6%/6%). Sequence lengths range from 5 to 287. **US Tornado:** The dataset is partitioned into 195 training sequences, 24 validation sequences, and 36 test sequences (train/validation/test: 76%/9%/14%). Sequence lengths range from 5 to 150. **eMAS:** The dataset is partitioned into 214 training sequences, 46 validation sequences, and 47 test sequences (train/validation/test: 70%/15%/15%). Sequence lengths are fixed at 50. **Crime:** The dataset is partitioned into 78 training sequences, 16 validation sequences, and 18 test sequences (train/validation/test: 70%/14%/16%). Sequence lengths range from 3 to 69. **Citibike:** The dataset is partitioned into 2,440 training sequences, 300 validation sequences, and 320 test sequences (train/validation/test: 80%/10%/10%). Sequence lengths range from 14 to 204. **Traffic Collision:** The dataset is partitioned into 233 training sequences, 50 validation sequences, and 51 test sequences (train/validation/test: 70%/15%/15%). Sequence lengths range from 11 to 199. **Twitter:** The dataset is partitioned into 36 training sequences, 6 validation sequences, and 6 test sequences (train/validation/test: 75%/12%/12%). Sequence lengths range from 10 to 235.

Baselines. We compare against representative methods spanning spatial, temporal, and spatio-temporal point processes. On the spatial side, **Conditional KDE** [39, 58] estimates the next-event location non-parametrically, while **CNF** [4] and its time-varying extension **TVCNF** [3] use invertible continuous transformations to model complex spatial densities, with TVCNF additionally conditioning on time.

For temporal prediction we cover both classical and neural TPPs. The **Poisson** process serves as a constant-rate reference; the **Hawkes** process [17] captures self-excitation, where past events raise future intensity; and the **Self-correcting** process [20] captures the opposite, where past events

suppress it. Among neural TPPs, **RMTPP** [12] encodes event history with an RNN, **NHP** [30] replaces the RNN with continuous-time LSTM dynamics, and **AttNHP** [31], **SAHP** [55], and **THP** [60] progressively replace recurrence with attention and Transformer architectures to model longer-range dependencies.

For joint spatio-temporal prediction, **NJSDE** [21] models events with latent jump SDEs, and **NSTPP** [3] combines neural ODE dynamics with normalizing flows over space. **DeepSTPP** [59] pairs neural sequence modeling with a learned spatial intensity, while **DSTPP** [54] and **TEN-DM** [27] cast next-event prediction as a diffusion process — the latter further injecting topological structure into the generative model.

Implementation Details: The optimization is performed using AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.99$), with a linear warm-up of the learning rate from 0 to 1e-3 over the first 5 epochs, followed by a linear decay to 5e-5 across 300 epochs. The STL and MSTL are trained with a reduced learning rate set to one-tenth of the base rate to ensure stable optimization. The diffusion model is configured with 500 timesteps under a cosine noise schedule. For the VLM-based encoder, we employ the Qwen2-VL [48]. Table 10 summarizes the dataset-specific hyperparameter configurations of MCM-DM. We keep the majority of hyperparameters fixed across datasets, and only adjust a small set of components to account for dataset-specific characteristics.

Diffusion for Spatio-Temporal Point Processes (STPPs). STPPs characterize asynchronous events through a conditional intensity function over time and space. Classical formulations include (in)homogeneous Poisson processes [9], the Hawkes process [18], Neyman-Scott processes [35], and Cox processes [8, 10], all of which rely on parametric kernels with limited expressiveness. To capture richer history dependencies, neural temporal point processes have been proposed [12, 30, 60, 41], and on the spatial side, Continuous Normalizing Flows (CNF) and their time-varying extension (TVCNF) parameterize event-location distributions [4, 3]. Influence-kernel-based deep approaches [11] further relax stationarity.

More recent neural STPPs introduce stronger structural inductive biases: NJDTPP [56] unifies classical TPPs as neural jump-diffusion SDEs, Decoupled MTPP [44] disentangles per-event latent trajectories with neural ODEs, and GNN-based kernels [2, 6] learn graph-structured event influence. Several surveys consolidate the field [42, 34, 6].

Generative formulations of STPPs have advanced rapidly, particularly diffusion-based approaches [54, 28, 40]. Yet despite their strong sample quality and natural support for uncertainty quantification, existing diffusion methods for STPPs operate solely on numerical timestamps and coordinates, ignoring the geographic semantics of where events actually occur.

Table 10: Hyperparameter Settings for MCM-DM.

Hyperparameter	**JPN Earthquake**	**COVID-19**	**US Tornado**	**eMAS**	**Crime**	**Citibike**	**Traffic Collision**	**Twitter**
Dropout	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Optimizer	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW
Embedding Dim / Heads	[64, 4]	[64, 4]	[64, 4]	[64, 4]	[64, 4]	[64, 4]	[64, 4]	[64, 4]
Transformer Layers	2	2	2	2	2	2	2	2
Satellite Image Zoom Level	12	12	7	7	12	15	12	14
Batch Size	64	64	64	64	64	64	64	64
Learning Rate (Diffusion)	1e-3	1e-3	1e-3	1e-3	1e-3	1e-3	1e-3	1e-3
Learning Rate (Morse/Gate/Encoder)	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4	1e-4
Diffusion Timesteps	500	500	500	500	500	500	500	500
Noise Schedule	cosine	cosine	cosine	cosine	cosine	cosine	cosine	cosine
η_1	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
η_2	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
η_3	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
Training Epochs	300	300	300	300	300	300	300	300

I Computational Complexity

Table 11: Running time comparison per epoch (in seconds (s)).

Model	**JPN Earthquake**	**COVID-19**	**US Tornado**	**eMAS**	**Crime**	**Citibike**	**Traffic Collision**	**Twitter**
DSTPP	44.835 s	56.542 s	29.138 s	29.227	25.162 s	86.796 s	36.589 s	16.783 s
TEN-DM	14.096 s	22.873 s	5.351 s	3.586 s	2.518 s	44.847 s	5.603 s	2.599 s
MCM-DM (Ours)	9.930 s	15.979 s	2.343 s	1.809 s	1.083 s	31.336 s	3.915 s	1.118 s

We analyse the time and space complexity of constructing the event skeleton $\mathrm{Skel}(\mathcal{S}) = (E_{cc}, E_{cp}, E_{pc}, E_{pp})$ on an STPP realisation $\mathcal{S} = \{x_1, \dots, x_L\}$ with $x_i = (t_i, s_i) \in \mathbb{R} \times \mathbb{R}^d$. The Morse function is treated as a free parameter; we give the bound for an arbitrary radially symmetric kernel and then specialise to the Matérn family. Throughout, $k = \lceil \sqrt{L} \rceil$ is the k-NN parameter, $d = O(1)$ is the spatial dimension, and after symmetrisation the graph $G = (V, E)$ has $|V| = L$ and $|E| \le 2Lk = O(Lk)$. We write τ_K for the per-evaluation cost of the kernel: $\tau_K = O(1)$ with a small constant for closed-form kernels (Gaussian, exponential, Matérn at half-integer ν), and $\tau_K = O(1)$ with a larger precision-dependent constant for general-ν Matérn via series and asymptotic expansions of the modified Bessel function K_ν.

I.1 Pipeline and per-stage costs

The pipeline has five stages. Stage 1 builds the k-NN graph on the joint metric $d(x_i, x_j) = \sqrt{(\Delta t)^2/\sigma_t^2 + \|\Delta s\|^2/\sigma_s^2}$. Using a KD-tree on the normalized points $(t_i/\sigma_t, s_i/\sigma_s) \in \mathbb{R}^{1+d}$, construction takes $O(L \log L)$ time and each k-NN query takes expected $O(k + \log L)$ time in fixed dimension under standard distributional assumptions [15, 1], for a total of $O(Lk + L \log L) = O(Lk)$ since $k = \lceil \sqrt{L} \rceil$ dominates $\log L$. Symmetrization is an additional $O(Lk)$ pass. This stage is shared by both Morse functions.

Stage 2 evaluates the spatial and causal-temporal Morse functions

$$g_s(x_i) \;=\; -\sum_{j \in \mathcal{N}(x_i)} K\Big(\tfrac{\|s_i - s_j\|}{\sigma_s}\Big) + \varepsilon_i^{(s)}, \qquad g_t(x_i) \;=\; -\sum_{\substack{j \in \mathcal{N}(x_i) \\ t_j < t_i}} K\Big(\tfrac{|t_i - t_j|}{\sigma_t}\Big) + \varepsilon_i^{(t)}, \tag{17}$$

where the kernel sum is restricted to the k-NN neighborhood $\mathcal{N}(x_i)$ of size at most $2k$ (the symmetrization factor). Each vertex contributes $O(k)$ kernel evaluations at cost τ_K each, so each Morse function costs $O(Lk \cdot \tau_K)$. The causal mask in g_t adds an $O(1)$ check per neighbor and does not change the asymptotic. Running both functions yields $T_2 = O(Lk \cdot \tau_K)$ with a constant of 2.

Stage 3 extends each vertex valuation to edges by the averaging rule $g(\{x_i, x_j\}) = \frac{1}{2}(g(x_i) + g(x_j))$ of [24], at $O(1)$ per edge and $O(Lk)$ total per Morse function, independent of τ_K. Stage 4 implements the greedy Joswig–Pfetsch ascending sweep [23]: vertices are sorted by g-value in $O(L \log L)$ time, and each is matched to its minimum-g free incident edge in $O(\deg(v))$ time, giving $\sum_v O(\deg(v)) = O(|E|) = O(Lk)$ across the sweep. Per Morse function the cost is $O(L \log L + Lk) = O(Lk)$, again independent of τ_K. Stage 5 computes the four-way intersection by hashing each edge to its (spatial-state, temporal-state) pair, at $O(|E|) = O(Lk)$ total.

Remark 10 (Role of Kannan averaging). The averaging extension guarantees the strict descending property $g(e_0) > g(v_0) > g(e_1) > \cdots$ along V-paths, which renders the resulting gradient field acyclic by construction. Alternative extensions that lack this property generally require an explicit acyclicity-repair pass on the Hasse diagram, which empirically inflates the constant in Stage 4 and may, in adversarial cases, introduce an additional $O(k)$ factor. We do not formalize this trade-off here; it is one practical reason to prefer the averaging rule.

I.2 Total complexity

For any radially symmetric kernel K with per-evaluation cost τ_K, the time complexity of constructing $\mathrm{Skel}(\mathcal{S})$ is

$$T_{\mathrm{total}}(L; K) \;=\; O\big(Lk \log L + Lk \cdot \tau_K\big) \;=\; O\big(L^{3/2}\, \tau_K\big),$$

with $k = \lceil \sqrt{L} \rceil$. Space complexity is $O(Lk) = O(L^{3/2})$, dominated by the symmetrised edge list and free/paired flags. In particular, stages 1, 3, 4, and 5 are each $O(Lk)$ and kernel-independent. Stage 2 contributes $O(Lk \cdot \tau_K)$. Substituting $k = \lceil \sqrt{L} \rceil$ gives the stated bound, with Stage 2 dominating whenever $\tau_K = \omega(1)$.

Remark 11 (Matérn specialisations). For $K = K_\nu$: (i) At closed-form smoothness $\nu \in \{1/2, 3/2, 5/2, \infty\}$, $\tau_{K_\nu} = O(1)$ with a small constant (one or two evaluations of $\exp$), and the total complexity is $O(L^{3/2})$. (ii) For general $\nu > 0$, evaluation of K_ν via series or asymptotic expansion remains $O(1)$ in L, with a constant that depends on requested precision and is, in typical double-precision implementations, roughly an order of magnitude larger than the Gaussian; the total complexity is still $O(L^{3/2})$.

The smoothness parameter ν is therefore asymptotically free in the Matérn family. The recommendation to use $\nu \in \{1/2, 3/2, 5/2, \infty\}$ is driven by the multiplicative constant in Stage 2, not by an asymptotic penalty.

Stage	Operation	Cost	Kernel-dependent
1	k-NN graph build (KD-tree)	$O(Lk)$	no
2	Vertex Morse values g_s, g_t	$O(Lk \cdot \tau_K)$	yes
3	Kannan edge averaging	$O(Lk)$	no
4	Greedy Forman pairing ($\times 2$)	$O(Lk + L \log L)$	no
5	Four-way intersection	$O(Lk)$	no
	Total	$O(L^{3/2}\, \tau_K)$	via Stage 2

I.3 Caveats and refinements

The k-NN restriction is load-bearing. Stages 2–5 depend on $|E| = O(Lk)$ rather than $O(L^2)$; this relies on K decaying fast enough that the truncated kernel sum in (17) is a faithful approximation to the full KDE. For Gaussian and Matérn with $\nu \geq 3/2$ the tail is super-polynomial and truncation is benign; for the heavy-tailed exponential case ($\nu = 1/2$) one may need a larger truncation radius or an explicit far-field correction, in which case the effective k in the analysis grows accordingly. Without truncation, Stage 2 alone costs $O(L^2 \tau_K)$ and the $O(L^{3/2})$ total no longer holds.

Constants for the four-way decomposition. Running both g_s and g_t rather than a single joint Morse function multiplies the constant in front of $L^{3/2}$ by approximately 2, as it requires two independent Stage 2 sweeps and two greedy pairings, against one shared k-NN graph and one intersection step.

J Broader Impact

This work advances spatio-temporal point process modeling by integrating geographic semantics and topological structure, which has the potential to benefit a wide range of applications such as disaster response, public health monitoring, urban planning, and environmental analysis. By improving the accuracy and interpretability of event forecasting, the proposed framework can support more informed decision-making in high-impact domains.

K Limitations

The proposed method may be less reliable in settings where geographic information is weak, noisy, or unavailable (e.g., low-quality satellite imagery or domains where spatial context is less informative). Future work could focus on improving robustness to missing modalities and exploring more lightweight or adaptive designs, as well as evaluating performance across a broader range of domains and real-world deployment scenarios.

L The Use of LLMs

Large language models (LLMs) were only used for rephrasing sentences and correcting grammar. Note that LLMs were not used for designing experiments, generating results, or contributing to the scientific content of the paper.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the paper's contributions and scope by clearly presenting the MCM-DM framework (VLM-based encoder, spatio-temporal fusion, and Morse-theoretic aligner), which are fully developed in the methodology and supported by theoretical analysis and experiments on multiple datasets.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The proposed method may be less reliable in settings where geographic information is weak, noisy, or unavailable (e.g., low-quality satellite imagery or domains where spatial context is less informative). Future work could focus on improving robustness to missing modalities and exploring more lightweight or adaptive designs, as well as evaluating performance across a broader range of domains and real-world deployment scenarios (see Appendix K).

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: This paper provides formal statements for its key theoretical results, including clearly defined assumptions (e.g., no-ties condition, structural margins, perturbation bounds) and a main stability theorem for the event skeleton, along with supporting propositions and detailed derivations in the appendix. The proofs are mathematically consistent and sufficiently rigorous for the claims made.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides substantial details for reproducibility, including dataset descriptions, evaluation metrics, baseline comparisons, and key implementation aspects (e.g., implementation details and hardware), and also includes an anonymous code repository to facilitate replication.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper provides open access to an anonymized code repository, enabling reproducibility of the main experimental results, and the appendix includes additional implementation details (e.g., datasets, training setup, and evaluation protocols) to support replication of the experiments.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies training and testing details, including dataset splits, evaluation metrics, hyperparameters, and general implementation setup.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The paper provides error bars for the proposed model and baselines, and discusses p-values.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper provide information on the computer resources in Appendix 5.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper uses public data. Authors have reviewed the NeurIPS Code of Ethics and do not identify deviations from it.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.

- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: This work advances spatio-temporal point process modeling by integrating geographic semantics and topological structure, which has the potential to benefit a wide range of applications such as disaster response, public health monitoring, urban planning, and environmental analysis. By improving the accuracy and interpretability of event forecasting, the proposed framework can support more informed decision-making in high-impact domains (see Appendix J).

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: [N/A]

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper cites prior work, datasets, and baselines used throughout the study.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The paper credits prior work, datasets, and models used throughout, and introduces its own implementation with an anonymized code release for submission.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing experiments or new research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not conduct crowdsourcing or prospective human-subject experiments.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: LLMs were used only for wording refinement, including improvements to grammar, clarity, and conciseness (see Appendix L).

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.