Supplementary Methodology

Homicide Hotspots in Chicago: Disentangling Structural Risk from Contagion Dynamics

Technical Documentation & User Manual
Version 2.4.0
Publication Year: 2026

Prepared for Submission as Peer-Reviewed Supplementary Material

This document provides extended methodological details supporting the analyses reported in the manuscript "Homicide Hotspots in Chicago: Disentangling Structural Risk from Contagion Dynamics". The information presented here is omitted from the main text for reasons of space and accessibility, and is provided to ensure transparency, replicability, and technical completeness.

Contents

Appendix A — Model Specification

A1. Spatiotemporal Point Process Framework

A fundamental starting point in describing Hawkes processes is to distinguish between a counting process and a point process. A counting process N_t can be viewed as a stochastic function describing the cumulative number of events that have occurred up to a given time t. The trajectory of N_t is a step function that starts at $N_0=0$, takes only nonnegative integer values, and increases by one unit each time an event occurs. A point process, in contrast, does not represent cumulative counts but directly encodes the exact instants when events occur. The most basic way to define a point process is as a sequence of nonnegative random variables, where each T_i represents the time at which the ith event occurs. Thus, the point process is represented by the sequence $T_1, T_2, \ldots$ of event times.

The central element in both approaches (counting and point-based) is the conditional intensity, which measures the instantaneous rate of occurrence of events at time t given the history H_t. The Poisson process provides the simplest reference point for specifying the conditional intensity in point processes. In a homogeneous Poisson process, events occur independently and with a constant rate over time. This implies that events are completely random—they do not depend on previous occurrences, and the probability of observing an event within an interval depends only on the interval's length, not its temporal position. However, if the rate varies deterministically with time, one obtains a non-homogeneous Poisson process. In this case, the conditional intensity remains independent of event history but reflects predictable variations in the rate over time.

> ■ **DEFINITION: HAWKES PROCESS INTENSITY**
> *In Hawkes-type processes, the event history is explicitly considered, implying that the intensity is stochastic or random, rather than deterministic. The probability of a new event rises following previous ones and gradually decays if no further events occur.*

By contrast, in Hawkes processes, the conditional intensity λ_t decomposes into two subcomponents: the background rate and the triggering or self-excitation function. Thus:

$$\lambda_t = \mu + \Sigma i : t_i < t \; \phi(t - t_i) \quad (1)$$

where μ denotes the background rate (exogenous events, independent of the event history) and $\phi(\cdot)$ represents the triggering function (endogenous component), which governs how each past event temporarily increases the future intensity. Consequently, the Hawkes process introduces historical dependence: each event raises the intensity function temporarily and/or spatially, thereby increasing the probability of new events in the short term and nearby in space. In other words, in Hawkes-type processes, the event history is explicitly considered, implying that the intensity is stochastic or random, rather than deterministic. The probability of a new event rises following previous ones and gradually decays if no further events occur.

A2. Specification of the Triggering Function

The triggering function can be specified in multiple ways, broadly categorized into parametric, nonparametric, and semiparametric forms. Among parametric specifications, the exponential form is widely used due to its simplicity and interpretability. This exponential triggering kernel assumes that the influence of an event on subsequent occurrences diminishes rapidly and at a constant rate. The simple exponential form is expressed as:

$$\phi(t) = \alpha e^{-\beta t} \qquad (2)$$

where α represents the magnitude or strength of excitation (how much the intensity of future events increases immediately after an occurrence), and β denotes the decay rate of the influence of the event history. A high value of α implies that an event has a greater capacity to generate subsequent events. The speed at which the historical influence dissipates depends on the magnitude of β. When β takes on a large value, the self-excitation effect fades rapidly.

Although the single exponential kernel is the most common, other specifications can capture more complex dynamics, such as a sum of exponentials (mixture kernel):

$$\phi(t) = \Sigma_{k=1}^{K} \alpha_k e^{-\beta k t} \qquad (3)$$

where K denotes the number of exponentials in the sum. When K=1, the specification reduces to a simple exponential kernel. The sum of exponentials is useful for modeling events that exhibit different temporal behaviors simultaneously. When employing an exponential kernel, it is possible to quantify the contribution of past events to new occurrences through the integral of the kernel, yielding the branching ratio or reproduction rate:

$$\Vert \phi \Vert_1 = \int_0^{\infty} \phi(t)dt = \alpha / \beta \qquad (4)$$

The result can be interpreted as the expected number of secondary events induced by each previous primary event. The second group of triggering function estimations corresponds to the non-parametric type. This approach allows the data themselves to determine the functional form of the process (Bacry & Muzy 2014; Bonnet & Sangnier 2024), rather than assuming specific predefined formulas that impose rigid structural assumptions and restrict the model's expressiveness (D'Angelo et al. 2024). Consequently, the choice between parametric and non-parametric modelling depends on the nature of the process and the availability of data.

Nevertheless, the literature applying Hawkes-type processes to the study of crime appears to show a preference for non-parametric and semi-parametric estimation approaches (Rosser & Cheng 2019; Mohler 2014; Mohler et al. 2011). This trend may be due to the limited number of empirical studies that provide insight into the spatio-temporal nature of crime data. In the absence of a robust empirical foundation, it is common to avoid assuming any specific spatio-temporal structure of the phenomenon under study, but rather to allow the data themselves to reveal it (Zhuang & Mateu 2019).

> **■ BEST PRACTICE: NONPARAMETRIC ESTIMATION**
>
> *Nonparametric approaches allow the data themselves to determine the functional form of the process, avoiding rigid structural assumptions that restrict model expressiveness. Quantitative criminology literature shows a strong emphasis on these data-driven structures to uncover spatio-temporal dynamics without imposing arbitrary structural bounds.*

A3. Background Rate and Risk Heterogeneity

In the simplest linear form of a unidimensional Hawkes process, μ is typically specified as a homogeneous constant, implying that the baseline density of events does not vary across space or time. However, empirical evidence consistently demonstrates that crime is tightly concentrated in particular locations and at specific times, emphasizing that the baseline rate must reflect this spatial concentration. When there is no clear evidence about which exogenous factors explain the concentration of crime, log-Gaussian Cox processes provide a flexible alternative for capturing latent heterogeneity. These models are considered suitable for capturing environmental heterogeneity and have been employed to analyze the socioeconomic conditions underlying crime clustering.

Process	Mechanism	Conditional Intensity
Hawkes	Models crime clustering by considering self-excitation. Emphasizes the importance of past event history. Linked to event dependence in criminological theory.	The intensity is stochastic and explicitly depends on past events. Dual component: background function and triggering function. The conditional intensity results from the sum of both components.
Cox Log-Gaussian	Models crime clustering based on unobserved heterogeneity or environmental factors. Linked to risk heterogeneity in criminological theory.	Dual stochastic process. The intensity is typically modeled by summing deterministic components (spatial and/or temporal covariates) and a Gaussian random field. Allows the incorporation of covariates into the model and the inclusion of a stochastic component.

Table A1: Comparative overview of Hawkes and Cox Log-Gaussian point process mechanisms.

A4. Spatio-temporal Cox-Hawkes Model

The analyses reported in the manuscript rely on a spatio-temporal Cox-Hawkes point process model, which combines a log-Gaussian Cox process for the background rate with a Hawkes process for self-excitation. The conditional intensity is defined as:

$$\lambda(s,t) = exp(a_0 + X_s^T w + f_s(s) + f_t(\mathbf{t})) + \mathbf{\Sigma} i : t_i < t\ \alpha\ f(t - t_i; \beta)\ \phi(s - s_i; \sigma) \quad (5)$$

The first term represents the log-Gaussian Cox component, where a_0 is the intercept, X_s denotes spatial covariates with associated weights w, and $f_s(s)$ and $f_t(t)$ are spatial and temporal Gaussian processes capturing latent background heterogeneity.

The second term represents the Hawkes component, where α denotes the reproduction rate, $f(t-t_i;\beta)$ is the temporal excitation kernel, and $\phi(s-s_i;\sigma)$ is the spatial excitation kernel governing the decay of influence over space. The temporal triggering function is specified as an exponential kernel (Equation 2), while the spatial kernel is Gaussian and centered on past events, implying that excitation effects are strongest near the original event location and decay smoothly with distance.

A5. Prior Specification

Bayesian inference requires the specification of prior distributions for all model parameters. Weakly informative priors were adopted following established architectural frameworks:

- For the parameter α: Beta(20,60). This assumes a prior mean value of 20/(20+60)=0.25, modeling the assumption that each event generates a mean of 0.25 new events under tight variance parameters (Var≈0.0023).
- For the parameter β (temporal decay): HalfNormal(2.0). This translates to an approximate mean of 2·2/π≈1.6, enforcing strictly positive support for the decay rate.
- For the parameter σ (spatial decay): HalfNormal(0.25). This yields an approximate prior spatial radius mean of 0.25·2/π≈0.20, restricting the immediate contagion radius to localized boundaries.

Gaussian processes in the log-Gaussian Cox component require hyperpriors for length scale and variance. To reduce computational burden, these priors are approximated using a pre-trained Variational Autoencoder.

A6. Inference Procedure

The Bayesian inference engine is implemented using the BSTPP framework developed by Manring et al. (2025). The analytical framework of this study relies completely on the underlying statistical mechanisms provided by BSTPP to perform inference via Stochastic Variational Inference (SVI). SVI approximates the posterior distribution using a parametric guide distribution and optimizes its parameters by maximizing the Evidence Lower Bound (ELBO). This approach allows efficient inference on large datasets with minimal loss of accuracy compared to Markov Chain Monte Carlo (MCMC) sampling.

■ **RECOMMENDATION: PIPELINE INTEGRATION**

The QUAKECRIME pipeline provides the analytical workflow, preprocessing procedures, validation tools, visualization modules, and reproducibility support without modifying the underlying Bayesian estimation algorithm or serving as an autonomous statistical estimator. Convergence is evaluated directly through the inspection of the SVI optimization history and the stabilization of the ELBO curve.

Appendix B — Model Validation and Predictive Performance

B1. Temporal Hold-Out Validation

To evaluate the predictive performance of the spatio-temporal Cox-Hawkes model without introducing artificial data leakage, a rigorous temporal hold-out validation framework was implemented.

- **Training Period:** The model parameters and latent spatial background fields are estimated exclusively on events falling within a closed historical window $[T_{start}, T_{end}]$.
- **Testing Period:** The validation metrics are evaluated on an out-of-sample forward-looking temporal window $[T_{sim\_start}, T_{sim\_end}]$.
- **Out-of-Sample Prediction:** Predictive surfaces are generated using only the structural parameters derived from the training window, ensuring that the model is tested on entirely unseen event patterns.

> ■ **IMPORTANT: METHODOLOGICAL CONSTRAINT**
> *Temporal hold-out validation is strongly preferred over random cross-validation splits for point process models. Randomly splitting point data breaks the intrinsic directional dependency of self-exciting processes, creating an artificial scenario where subsequent 'training' events appear to trigger historical 'testing' events. Temporal partitioning maintains the causal history Ht, providing an honest test of real-world forecasting utility.*

B2. Predictive Performance Metrics

The out-of-sample risk surfaces were evaluated against the spatial coordinates of held-out validation events mapped onto an evaluation raster grid.

ROC Curve and Area Under the ROC Curve (ROC AUC)

The Receiver Operating Characteristic (ROC) curve plots the True Positive Rate against the False Positive Rate across all possible probability thresholds. The ROC AUC summarizes this curve:

AUC = $\int_0^1 TPR(FPR)\, dFPR$

• **Interpretation:** Measures the probability that a randomly chosen occupied cell is assigned a higher risk value than a randomly chosen unoccupied cell.
• **Advantages:** Scale-invariant and threshold-independent discrimination metric.
• **Limitations:** Highly susceptible to over-optimism under extreme spatial class imbalance (where the vast majority of grid cells are unoccupied).
• **Appropriateness for Homicide Prediction:** Useful as a baseline measure of broad spatial discrimination between safe and high-risk environments.

Precision–Recall Curve and Average Precision (AP)

The Precision–Recall curve balances positive predictive value against sensitivity. Average Precision summarizes this trade-off:

$\mathbf{AP} = \Sigma n\,(R_n - R_{n-1})\,P_n$

where P_n and R_n are the precision and recall at the nth threshold.

- **Interpretation:** Quantifies the average fraction of correctly predicted crime locations among all cells flagged as high risk.
- **Advantages:** Focuses directly on the minority class (occupied cells), providing an accurate measure of predictive utility in sparse spatial data.
- **Limitations:** Highly dependent on the baseline prevalence of events within the grid layout.
- **Appropriateness for Homicide Prediction:** Crucial for criminological applications, as it directly reflects the operational efficiency of targeted tactical deployments.

Brier Score

The Brier Score measures the mean squared error of the probabilistic forecasts:

$BS = (1 / M)\,\Sigma_{m=1}^{M}\,(p_m - y_m)^2$

where p_m is the count-calibrated Poisson occupancy probability (p_m=1−exp(−λ_m)) for cell m, $y_m \in \{0,1\}$ is the empirical binary occupancy status, and M is the total number of evaluation cells.

- **Interpretation:** Quantifies the calibration and accuracy of the localized probability assignments.
- **Advantages:** Strictly proper scoring rule that penalizes both overconfidence and underconfidence.
- **Limitations:** Dominated by the zero-class score when spatial crime patterns are highly sparse.
- **Appropriateness for Homicide Prediction:** Ensures that the magnitude of predicted risk tracks the true probability of violence.

B3. Tactical Hotspot Metrics

Criminological utility is assessed via operational metrics defined over specific proportions of the spatial study domain.

Predictive Accuracy Index (PAI)

Formally defined as the ratio of the hit rate to the area coverage:

$PAI = \text{Hit Rate} / \text{Area Coverage} = (n_{\text{hits}} / N_{\text{total}}) / (a_{\text{hotspot}} / A_{\text{total}})$

- **Interpretation:** Measures the efficiency of the hotspot definition. A PAI >1 indicates that the model concentrates crime predictions more effectively than a random spatial allocation.
- **Criminological Value:** Widely used in hotspot policing to justify the spatial optimization of patrol boundaries, directly maximizing resource efficiency.

Hit Rate

The Hit Rate is the proportion of total out-of-sample validation events that fall within the designated hotspot boundary:

Hit Rate = $n_{\text{hits}} / N_{\text{total}}$

It indicates the raw sensitivity of the tactical boundaries, mapping the total volume of violent crime intercepted by the forecast.

Area Coverage

The fraction of the total active study domain selected as a high-risk hotspot:

Area Coverage = $a_{\text{hotspot}} / A_{\text{total}}$

Operationally, it reflects the resource footprint demanded from an agency to maintain full coverage across the designated target zones.

B4. Risk Ranking and Spatial Prediction Maps

To generate operational outputs, the continuous predicted background intensity surface λ(s,t) is evaluated across a dense spatial grid. Cells whose centers fall within the valid geographic domain are sorted in descending order of their predicted risk values. Hotspot boundaries are drawn by extracting the top k cells required to satisfy fixed area coverage thresholds (e.g., 1%, 5%, 10%, and 20%). The final spatial prediction maps display these nested risk tiers alongside out-of-sample validation points to visually assess the alignment between structural background risk and subsequent homicide occurrences.

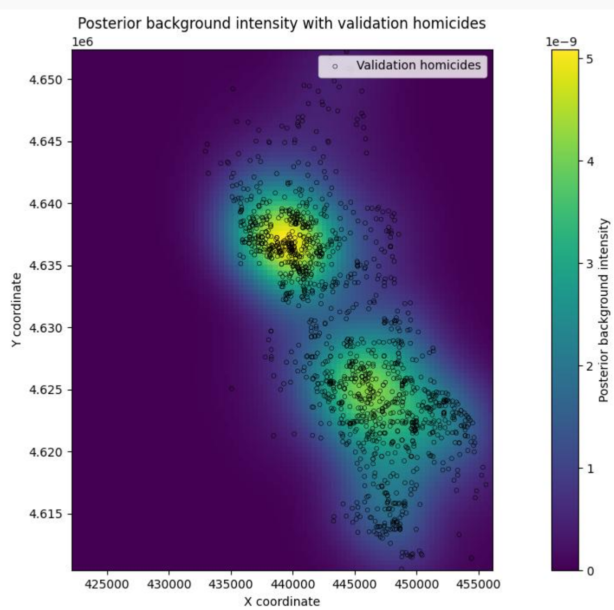

Figure B1: Graphical representation of nested tactical risk zones overlaying empirical evaluation data.

Appendix C — Robustness and Sensitivity Analyses

C1. Coordinate-System Comparison

Because point process models leverage exact spatial distance matrices, spatial inferences can be highly sensitive to coordinate representation. Controlled A/B/C tests were executed across three coordinate schemas within QUAKECRIME:

- **Schema A (Legacy Exact):** Employs original geographic coordinates (Long/Lat) mapped directly to an unprojected EPSG:4326 domain utilizing a 0.00015-degree street-network buffer.
- **Schema B (WGS84 Corrected):** Recalculates spatial positions directly from geometry centroids in EPSG:4326 using equivalent angular buffer tolerances.
- **Schema C (Projected Metric):** Transforms spatial inputs to an optimal estimated metric projection (UTM zone) utilizing a true 15-meter street-network buffer.

Comparing these structures evaluates whether choice of coordinate system shifts parameters like the spatial decay kernel (σ), ensuring that spatial conclusions reflect physical processes rather than projection distortions.

C2. Duplicate-Event Sensitivity

Administrative crime databases frequently contain duplicate entries at identical coordinates due to multi-victim incidents or repeated reporting. In a Hawkes self-exciting framework, identical coordinates create space-time pairs separated by exactly zero distance and time. This structural singularity can artificially drive the estimated triggering parameters toward zero-scale limits. QUAKECRIME executes automated pair diagnostics to isolate these occurrences and tests estimation stability by comparing models fitted with and without exact same-date, same-location analytical duplicates.

C3. Alternative Event Definitions

Criminological analyses can vary depending on whether the unit of analysis is defined as individual administrative victim records or unique police incidents. If a single homicide incident involves multiple victims, treating each victim as an independent event introduces multiple simultaneous points into the point process. Models were systematically compared using both configurations: (1) all administrative records, and (2) deduplicated incident-level entries (one event per unique Case_Number). This analysis is critical to verify that the estimated contagion dynamics reflect true sequential retaliation rather than administrative multi-victim clusters.

C4. Convergence and Stochastic Optimization Assessment

Stochastic Variational Inference relies on random mini-batch sampling and gradient descent, which can introduce instability depending on hyperparameter settings. To ensure structural reliability, optimization pathways were systematically monitored across varying step counts (2,000, 3,000, 5,000, 10,000, and 15,000 steps) under a fixed learning rate (η=0.02). Convergence diagnostics require that key trigger parameters (α, β, σ) stabilize asymptotically alongside the ELBO history curve. Substantive interpretation is deferred if parameter estimates continue to diverge at higher step intervals, isolating the impacts of stochastic optimization noise from the underlying data structure.