Spatial Point Patterns of Hawker Centres in Singapore

Alvin Bong

Abstract

Hawker centres are an essential part of Singapore's food landscape, providing affordable and culturally significant meals despite the country's high cost of living. This study examines the spatial distribution and accessibility of hawker centres across mainland Singapore. Using datasets from the Singapore Open Data Portal, we applied quadrat analysis, Ripley's K-function, kernel density estimation (KDE), and a Log-Gaussian Cox Process (LGCP) model fitted via INLA. Spatial randomness tests consistently indicate clustering, with concentrations most prominent around downtown Singapore. The LGCP model shows that proximity to MRT stations negatively correlates with hawker centre intensity, while population density shows a positive association. The spatial effects highlight additional unexplained clustering in central regions. Overall, this study demonstrates how spatial statistical methods can reveal accessibility patterns and structural drivers of Singapore's food infrastructure.

Introduction

Like many countries in the region, Singapore's food landscape plays a central role in many Southeast Asian countries, reflecting both cultural diversity and social functionality. Given that Singapore is a highly developed country with a relatively high cost of living, one might expect food to be uniformly expensive, as is often the case in many first-world nations. Yet Singapore stands out for offering a remarkably wide range of food prices, from inexpensive SGD $5 noodle dishes to premium fine-dining experiences. These affordable options are most commonly found in hawker centres, which are an iconic part of Singaporean culture. Hawker centres provide affordable, diverse, and convenient meals for the general population, especially for working adults with long hours who may not have the time or opportunity to cook at home.

Understanding the spatial distribution and accessibility of hawker centres provides insight into urban planning, food equity, and how infrastructure supports daily life. The main research objectives of this study are to:

- Examine accessibility of hawker centres through proximity to MRT stations, as a proxy for transit-oriented access.
- Assess whether the locations of hawker centres are spatially random or clustered, using quadrat counts and Ripley's K-function.
- Identify spatial clusters and intensity patterns through both non-parametric kernel density estimation (KDE) and a model-based approach using a Log-Gaussian Cox Process (LGCP), incorporating covariates such as population density and nearest distance to MRT stations.

These objectives allow us to quantitatively understand how accessibility, population, and urban design influence the spatial organization of Singapore's food infrastructure.

Data

The study focuses on mainland Singapore, omiting smaller outlying islands. All of the datasets (point coordinates of hawker centres, the boundary polygon and covariates) were obtained from the Singapore Open Data Portal (Land Transport Authority 2019; National Environment Agency 2019; Urban Redevelopment Authority 2019). Data were subsequently cleaned, processed and projected to EPSG:3414 (SVY21 / Singapore TM) to allow accurate distance measurements and spatial analysis.

Covariates

Proximity to MRT: MRT station locations were represented as points. For each hawker centre, the nearest proximity from each hawker centers was calculated as the minimum distance `st_distance()` to all MRT stations. This distance serves as a proxy for transit accessibility, reflecting how easy it is for residents and commuters to reach food services. Lower distances indicate higher accessibility, which may correlate with higher demand.

Population density: Population data were provided as areal values aggregated at the planning area level. Population density was calculated by dividing the total population by the area of the corresponding region using `st_area()`. This covariate serves as a proxy for potential demand: areas with higher population densities are expected to support more hawker centres, reflecting higher foot traffic and customer volume.

By including both MRT proximity and population density, the study captures two key structural drivers of hawker centre locations: accessibility and demand.

Methods

Spatial Randomness

Quadrat Method

The quadrat method divides the study region into 24 sub-regions and compares the observed number of hawker centres in each sub-region to the expected number under Complete Spatial Randomness (CSR).

$$E = \frac{\text{total count of hawker centres}}{\text{number of sub-regions}} = \frac{129}{24} \approx 5.4$$

Two approaches were applied, Chi-square test and Monte Carlo simulation. The former is suitable when expected coutns are reasonably large and windows are generally regular. The later generates CSR patterns to create a reference distribution of the test statistic, accommodating irregular windows or uneven counts per quadrat and is robust in Singapore's case, where sub-regions vary in shape and size. Both approaches compute the Pearson χ^2 statistic, but the p-value differs: the chi-square test uses the theoretical distribution, while the Monte Carlo test estimates the p-value empirically from simulations.

Figure 1: Point locations of 129 hawker centers in Singapore.

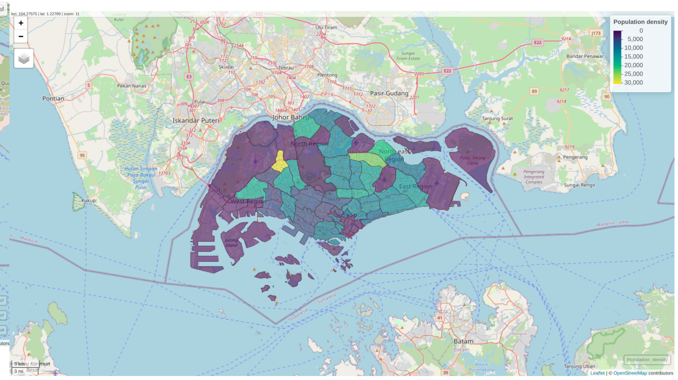

Figure 2: Population Density by planning area.

Figure 3: Point locations of MRT stations.

For chi-square test, the test statistic defined as:

$$X^2 = \sum_{i=1}^{24} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{24-1},$$

where O_i is the observed count and E_i is the expected count under CSR. This is used to conduct the hypothesis test:

$$H_0 : \text{Complete Spatial Randomness (CSR)}$$
$$H_1 : \text{Clustered pattern (departure from CSR)}$$

K-function

For any distance (s), the K-function is defined as:

$$K(s) = \lambda^{-1}\,\mathbb{E}\left[\text{number of further events within distance } s \text{ of an arbitrary event}\right],$$

where λ is the intensity of the spatial point process.

To assess deviation from CSR, we compare the empirical $K(s)$ to the theoretical K-function under a homogeneous Poisson process. $K_{\text{CSR}}(s) = \pi s^2$. - $K(s) > \pi s^2$ indicates **clustering** while $K(s) < \pi s^2$ indicates **inhibition/regularity**.

In practice, instead of relying solely on this theoretical expectation, Monte Carlo simulations of CSR are often used to generate an envelope (confidence interval) for $K(s)$ at each distance s. This accounts for stochastic variability in finite samples. In our case we set 99 simulations.

Intensity

Kernel Density Estimate (KDE)

KDE is used to estimate the intensity of points in space, providing a smoothed surface that highlights regions with dense hawker centres. This method allows identification of spatial clusters without assuming a specific underlying distribution. At location x, the estimated intensity is given by:

$$\hat{\lambda}(x) = \sum_{i=1}^{n} \frac{1}{h^2} K\Big(\frac{x - x_i}{h}\Big),$$

where h is the bandwidth and $K(\cdot)$ is the kernel.

Before applying KDE, areas where hawker centres cannot realistically occur, such as airports, water catchments, and mountainous regions, were excluded from the study region. This ensures that the estimated intensity surface accurately reflects feasible locations, avoids artificially low intensities in uninhabitable areas, and reduces edge effects during smoothing. A **Gaussian kernel** with bandwidth of **1200 meters** was used (via `spatstat::density()`). The resulting intensity surface visually identifies hotspots and clustering patterns, providing a descriptive baseline for comparison with model-based LGCP estimates.

Log-Gaussian Cox process model

We assume the spatial point pattern of hawker centres in Singapore $\{s_i : i = 1, \dots, n\}$ is generated by a **log-Gaussian Cox process** with intensity $\Lambda(s) = \exp(\eta(s))$. Following (Illian et al. 2012; Moraga 2020), the model is fitted by approximating the LGCP as a **latent Gaussian model** on a grid. The study region is discretized into an $n_1 \times n_2 = N$ grid of cells s_{ij}.

For cell s_{ij}, the mean number of points is:

$$\Lambda_{ij} = \int_{s_{ij}} \exp(\eta(s))\, ds \approx |s_{ij}| \exp(\eta_{ij}),$$

where $|s_{ij}|$ is the cell area. Then, conditional on the latent field, the observed count y_{ij} within each cell as follows:

$$y_{ij}|\eta_{ij} \sim Poisson(|sij| \cdot exp(\eta_{ij}))$$

$$\eta_{ij} = \mathbf{x}_{ij}^{\top}\beta + f_s(s_{ij}) + f_u(s_{ij})$$

where:

- β_0 is the intercept
- β_1, β_2 are regression coefficients for the covariates
- $f_s(s_{ij})$ is a spatially structured random effect reflecting unexplained variability that can be specified as a second-order two-dimensional CAR-model on a regular lattice
- $f_u(s_{ij})$ is a unstructured random effect reflecting independent variability in cell s_{ij}

Model fitting was performed in a Bayesian framework using the Integrated Nested Laplace Approximation (INLA) (Rue, Martino, and Chopin 2009).

Results

Exploratory Data Analysis

The nearest-MRT distances for hawker centre ranges from 59 meters to 2418 meters, with median and mean at 656 meters and 721 meters respectively.

Spatial Randomness

The test statistic of quadrat test for both ChiSquare and Monte Carlo approach is 113.4, with p-values < 0.05. Therefore, we conclude that the result is significant, and there is sufficient evidence to reject the null hypothesis H_0, suggesting that the locations of hawker centers in Singapore are clustered (not random).

Additionally, this is consistent with our k-function result (Figure 4), where it lies above the theoretical CSR envelope, indicating significant clustering across a range of up to 6km and more.

Applying the Kernel denstiy estimation (KDE) (Figure 5), it can be observed that the cluster of hawker centers is located towards the south, near to downtown Singapore.

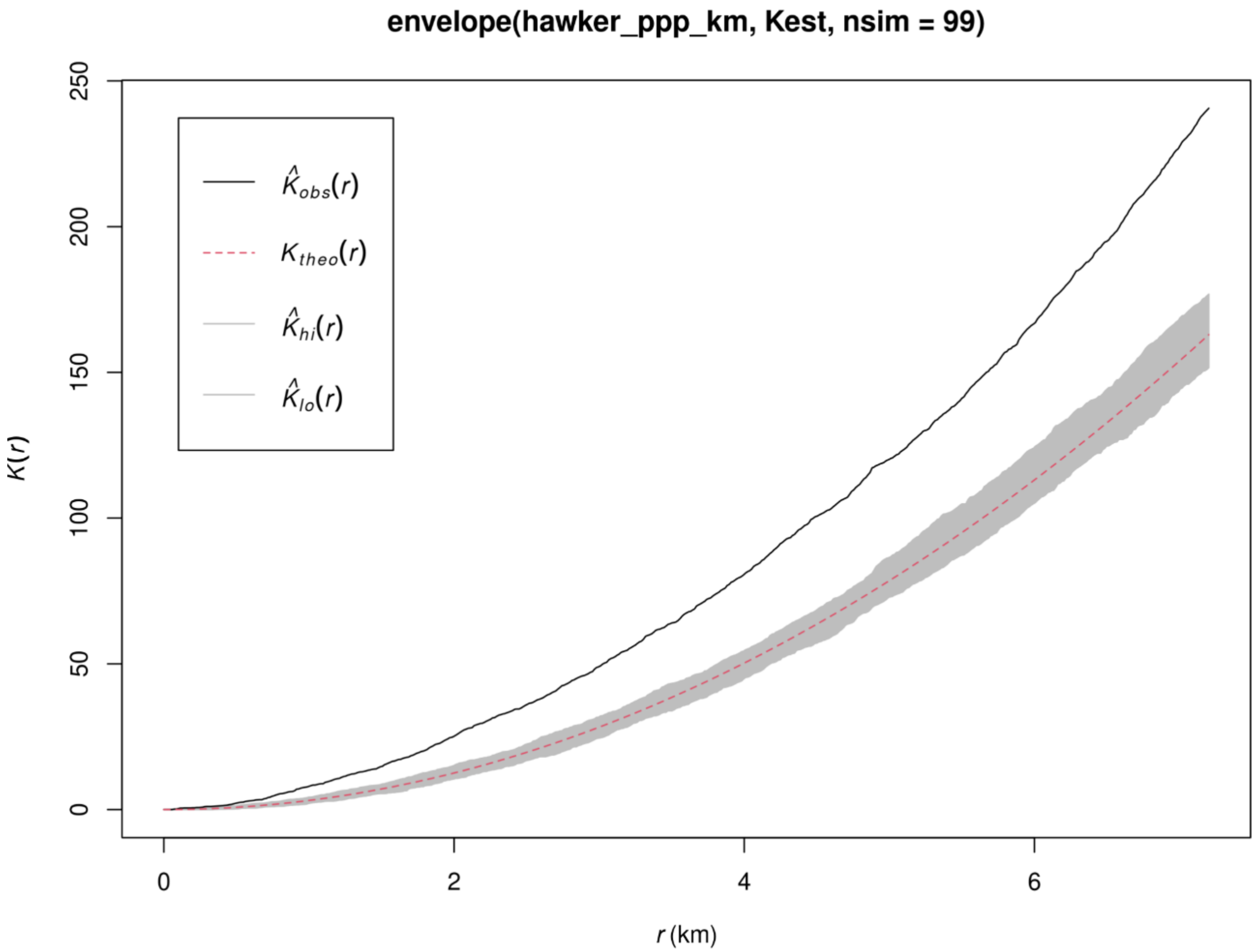

Figure 4

Model

The LGCP model estimated the intercept at $\hat{\beta}_0 = -14.5$ (95%~CI: -15.0, -14.0). Both covariates (nearest distance to MRT and population density) are significant. The coefficient for nearest distance to MRT is negative, $\hat{\beta}_1 = -0.0015$ (95%~CI: -0.0020, -0.0011), suggesting that there are fewer hawker centers further away from MRT stations. Conversely, the coefficient of population density is positive, $\hat{\beta}_2 = 4.5 \times 10^{-5}$ (95%~CI: -1.3×10^{-5}, 7.8×10^{-5}), suggesting hawker centers are abundant in areas with denser population. However, the posterior mean of unstructured random effect are not random, showing higher values in the south, suggesting there may be factors that have not been considered in the model.

Discussion

The spatial analysis demonstrates that hawker centres in Singapore are strategically located to balance accessibility and population demand. The median distance of roughly 656 meters suggests

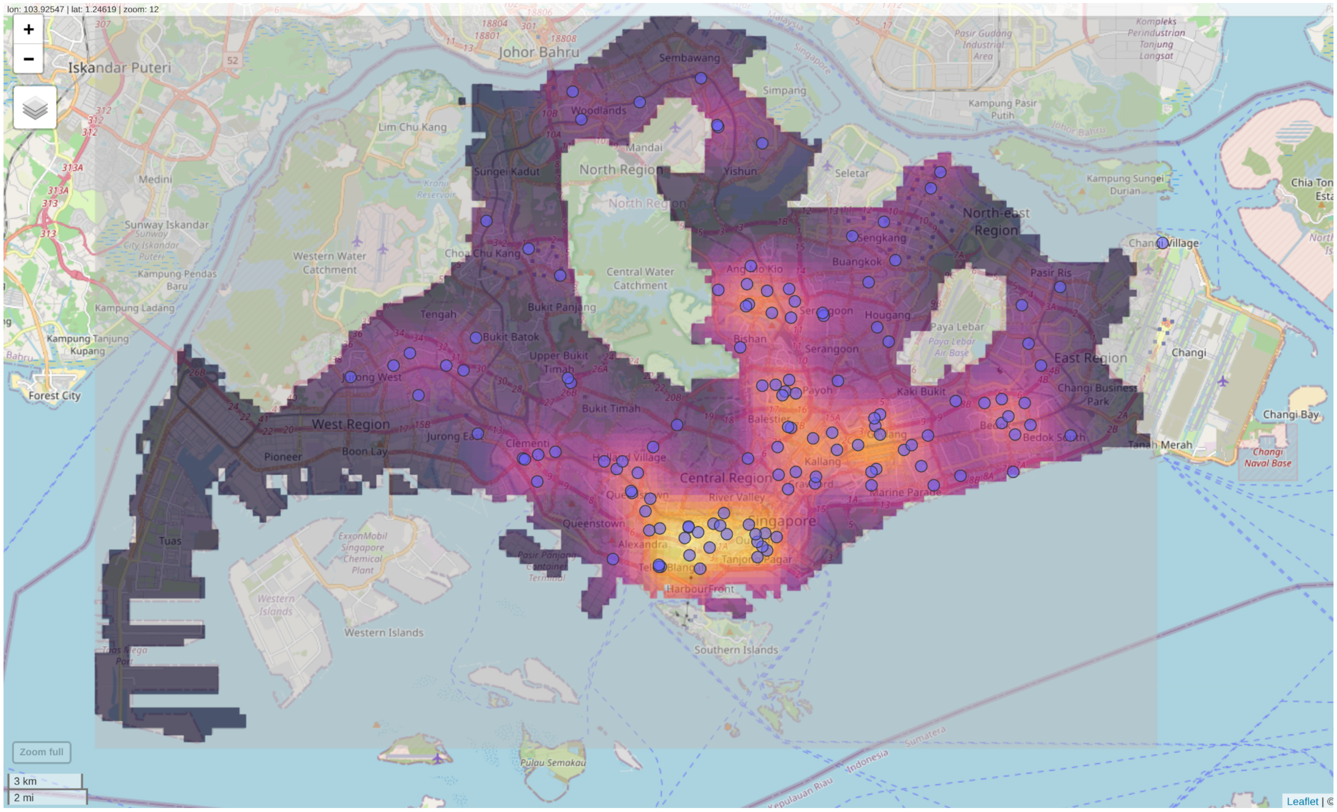

Figure 5: KDE higlighting high areas with dense hawker centers.

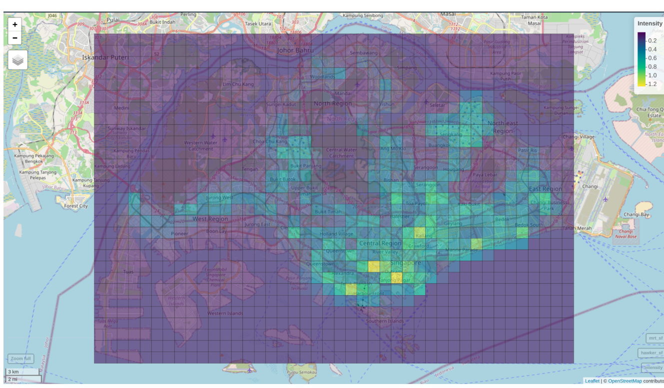

Figure 6: Predictions of mean intensity of hawker centers across Singapore

Figure 7: Posterior mean of unstructured random effect

that hawker centers in Singapore are generally located within close, walkable distance to MRT stations. Given that Singapore's MRT system is a critical component of daily mobility, this reflects urban planning that prioritizes accessibility and convenience for both residents and visitors. For commuters and workers, the proximity supports quick access to affordable meals.

Next, the quadrat analysis, K-function, and kernel density estimation consistently reject the assumption of complete spatial randomness and instead point to a clearly clustered spatial structure. The strongest intensities appear near downtown Singapore, indicating that hawker centers tend to aggregate around central commercial hubs. This clustering may be driven by both historical planning decisions and contemporary patterns of commuter movement, where central areas provide a steady demand for convenient and affordable meals. In other words, this may be a reflection of how food infrastructure evolves alongside population movements and economic activity.

The LGCP model further confirms these findings quantitatively: the intensity of hawker centres increases with population density and decreases with MRT distance, highlighting the importance of accessibility and human activity in shaping food infrastructure. The posterior mean intensity surface further identifies notable hotspots in Chinatown, Bukit Merah, Braddell, Joo Chiat, and Rochor.

Limitations

While the analysis provides meaningful insights, several limitations should be acknowledged.

First, both Ripley's K-function and kernel density estimation (KDE) are sensitive to edge effects, which can be particularly problematic in regions with irregular boundaries such as Singapore's coastline. Although standard edge-correction methods were applied, residual edge bias may still influence the apparent degree of clustering.

Second, the posterior mean of the unstructured random effect shows clear spatial structure, with higher values in the south. This indicates that relevant explanatory factors remain unaccounted for in the model. The set of covariates used was limited to MRT proximity and population density, but other potentially important drivers, such as housing type (HDB vs. private), commercial land-use intensity, socioeconomic characteristics, or daytime vs. nighttime population, were not included. The population density layer used was also static, despite significant daily mobility patterns in Singapore.

Finally, the spatially structured component was modelled using a second-order random walk (RW2) on a lattice. While appropriate, future work could explore more flexible spatial formulations such as SPDE-based models, which may better capture complex spatial dependencies.

Conclusions

In conclusion, this study demonstrates that hawker centres in Singapore exhibit clear spatial structure shaped by accessibility and population concentration. Exploratory analyses, including quadrat tests, Ripley's K-function, and KDE consistently reject spatial randomness and reveal pronounced clustering around the southern and central regions. The LGCP model further shows that hawker centre intensity increases with population density and decreases with distance from MRT stations, confirming the role of transit access and human activity in influencing food infrastructure. Remaining unexplained spatial patterns suggest that additional socio-economic and land-use factors may contribute to these distributions. Overall, this work highlights how spatial statistical methods can capture both planned and emergent aspects of Singapore's food landscape, offering a quantitative foundation for future studies and urban planning considerations.

References

Illian, Janine Barbel, Sigrunn Holbek Sørbye, Håvard Rue, and Ditte Katrine Hendrichsen. 2012. "Using INLA to Fit a Complex Point Process Model with Temporally Varying Effects-a Case Study." *Journal of Environmental Statistics* 3 (7).

Land Transport Authority. 2019. "LTA MRT Station Exit (GEOJSON) (2025)." https://data.gov.sg/datasets/d_b39d3a0871985372d7e1637193335da5/view.

Moraga, Paula. 2020. "Species Distribution Modeling Using Spatial Point Processes: A Case Study of Sloth Occurrence in Costa Rica."

National Environment Agency. 2019. "Hawker Centres (GEOJSON) (2025)." https://data.gov.sg/datasets/d_4a086da0a5553be1d89383cd90d07ecd/view.

Rue, Håvard, Sara Martino, and Nicolas Chopin. 2009. "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71 (2): 319–92.

Urban Redevelopment Authority. 2019. "Master Plan 2019 Subzone Boundary (No Sea) (GEOJSON) (2025)." https://data.gov.sg/datasets/d_8594ae9ff96d0c708bc2af633048edfb/view.