

# ***Technical Report***

## Chi-squared Cox Processes

Tilman, Charlotte (?), Adrian (?), others (?), ...

September 16, 2025

## 1 Introduction

Cox processes (doubly stochastic Poisson processes) are a flexible class of models for clustered spatial point patterns. The *log-Gaussian Cox process* (LGCP) is by far the most widely studied and applied, owing to its tractability and natural connection to Gaussian random fields (GRFs).

An alternative—rarely used in practice—is the *chi-squared Cox process* (CSCP), in which the random intensity is given by the sum of squared Gaussian fields. Despite their similarities, the two processes induce different distributional and clustering properties. This document summarizes key properties of both models, and sketches possible avenues for publication that highlight the merits and potential uses of CSCPs.

## 2 Definitions

### 2.1 Log-Gaussian Cox Process (LGCP)

An LGCP is defined by

$$\Lambda(s) = \exp\{Z(s)\}, \quad s \in D \subset \mathbb{R}^d,$$

where  $Z(s)$  is a Gaussian random field with mean function  $\mu(s)$  and covariance  $C(s, s')$ . Conditional on  $\Lambda$ , the point process  $X$  is Poisson with intensity function  $\Lambda(s)$ .

Key properties:

- $\mathbb{E}[\Lambda(s)] = \exp(\mu(s) + \frac{1}{2}\sigma^2)$ .
- The pair correlation function is

$$g(s, s') = \exp\{C(s, s')\},$$

showing that clustering strength depends exponentially on covariance. In particular,  $g(0) = \exp(\sigma^2)$  can be arbitrarily large.

- Widely used in practice: tractable through moment properties, simulation methods, and fitting algorithms (composite likelihood, Bayesian inference, INLA).

### 2.2 Chi-squared Cox Process (CSCP)

A CSCP is defined by

$$\Lambda(s) = \sum_{i=1}^k Z_i^2(s), \quad s \in D,$$

where  $\{Z_i(s)\}_{i=1}^k$  are independent Gaussian random fields with mean  $\mu_i(s)$  and covariance functions  $C_i(s, s')$ .

Key properties:

- $\mathbb{E}[\Lambda(s)] = \sum_{i=1}^k (\mu_i(s)^2 + \sigma_i^2)$ .
- $\text{Var}[\Lambda(s)] = 2 \sum_{i=1}^k (\sigma_i^4 + 2\mu_i^2\sigma_i^2)$ .
- The pair correlation has the form

$$g(s, s') = 1 + \frac{\sum_{i=1}^k (\text{Cov}(Z_i(s), Z_i(s')))^2}{(\sum_{i=1}^k \sigma_i^2 + \mu_i^2)^2}.$$

In the symmetric zero-mean case,  $g(0) = 1 + 2/k$ , so clustering strength at very short lags is bounded.

- Marginally,  $\Lambda(s)$  follows a noncentral chi-squared distribution with  $k$  degrees of freedom and noncentrality parameters  $\mu_i^2/\sigma_i^2$ . These marginals have lighter (exponential) tails compared to the lognormal marginals of an LGCP.

### 3 Similarities and Differences

#### Similarities

- Both are Cox processes driven by latent Gaussian random fields.
- Both induce clustering via spatial correlation in the latent fields.
- Both can represent multi-scale clustering when multiple latent fields with different correlation lengths are included.
- Both are second-order intensity reweighted stationary under suitable conditions, allowing use of summary statistics such as the pair correlation function.

#### Differences

- **Marginals:** LGCP intensities are lognormal with heavy tails; CSCP intensities are chi-squared/gamma-like with lighter exponential tails.
- **Pair correlation:** For LGCP,  $g(s, s') = \exp\{C(s, s')\}$  and  $g(0)$  can be arbitrarily large; for CSCP,  $g(s, s')$  depends quadratically on covariance and  $g(0)$  is bounded, yielding less explosive but potentially more interpretable clustering strength.
- **Role of  $k$ :** The degrees of freedom  $k$  in CSCP controls variability and relative clustering. For large  $k$ , the CSCP intensity approaches Gaussianity by the central limit theorem, reducing to near-homogeneous Poisson behavior.
- **Interpretability:** In LGCP, the log-scale linear predictor is directly interpretable. In CSCP,  $k$  may be viewed as the number of independent clustering mechanisms.

See Figure 1 for six realisations of a zero-mean, exponential-covariance CSCP with  $k = 2$ .

### 4 Covariance structure of the CSCP intensity field

Let the chi-squared Cox process (CSCP) be driven by  $k$  independent Gaussian random fields

$$Z_i(s) \sim \text{GRF}(\mu_i, \sigma_i^2 \rho_i(\|s - s'\|)), \quad i = 1, \dots, k,$$

where  $\mu_i$  is the mean,  $\sigma_i^2$  the marginal variance, and  $\rho_i(h)$  the correlation function at spatial lag  $h = \|s - s'\|$ . The random intensity is

$$\Lambda(s) = \sum_{i=1}^k Z_i(s)^2.$$

$k = 2$ ;  $\mu = 0, 0$ ;  $\text{scale} = 0.02, 0.2$ ;  $\text{var} = 500, 500$

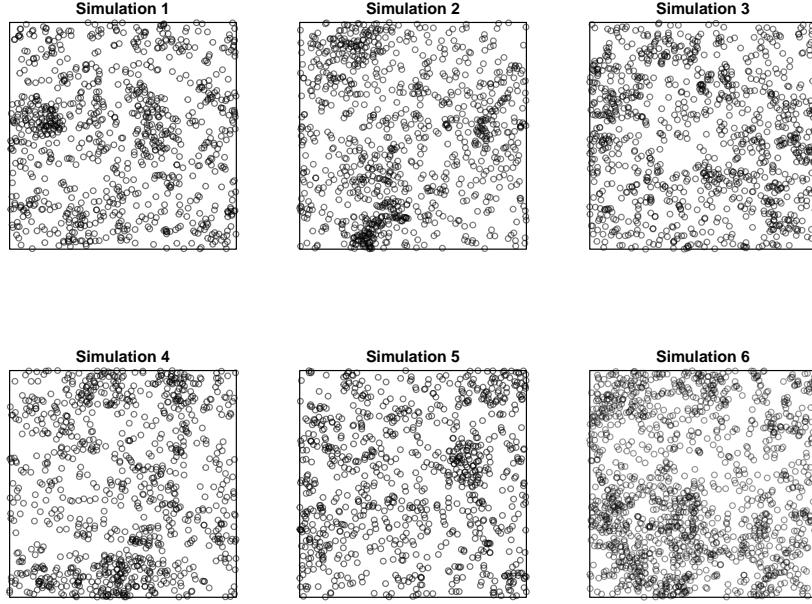


Figure 1: Six replicates of a 2-component CSCP.

#### 4.1 Mean intensity

By independence across  $i$ ,

$$\mathbb{E}\{\Lambda(s)\} = \sum_{i=1}^k \mathbb{E}\{Z_i(s)^2\} = \sum_{i=1}^k (\mu_i^2 + \sigma_i^2).$$

#### 4.2 Covariance function (general case)

For any two locations  $s, s'$  with separation  $h = \|s - s'\|$ ,

$$\text{Cov}_\Lambda(h) = \text{Cov}(\Lambda(s), \Lambda(s')) = \sum_{i=1}^k \text{Cov}(Z_i(s)^2, Z_i(s')^2),$$

since the fields are independent across  $i$ .

Applying Isserlis' (Wick's) theorem for fourth moments of Gaussians,

$$\mathbb{E}[Z_i(s)^2 Z_i(s')^2] = \mathbb{E}[Z_i(s)^2] \mathbb{E}[Z_i(s')^2] + 2 (\text{Cov}(Z_i(s), Z_i(s')))^2.$$

Thus

$$\text{Cov}(Z_i(s)^2, Z_i(s')^2) = 2(\sigma_i^2 \rho_i(h))^2 + 4\mu_i^2 \sigma_i^2 \rho_i(h).$$

Hence the full covariance is

$$\text{Cov}_\Lambda(h) = \sum_{i=1}^k [2\sigma_i^4 \rho_i(h)^2 + 4\mu_i^2 \sigma_i^2 \rho_i(h)].$$

### 4.3 Zero-mean special case

If  $\mu_i = 0$  for all  $i$ , this reduces to

$$\text{Cov}_\Lambda(h) = 2 \sum_{i=1}^k \sigma_i^4 \rho_i(h)^2.$$

This case is often taken as the baseline in the literature (e.g. Møller and Waagepetersen, 2004 [1]).

### 4.4 Pair correlation function

The pair correlation function is defined as

$$g(h) = 1 + \frac{\text{Cov}_\Lambda(h)}{(\mathbb{E}\{\Lambda(s)\})^2}.$$

For the zero-mean, equal-variance case with exponential correlation  $\rho_i(h) = \exp(-h/\phi_i)$ ,

$$g(h) - 1 = \sum_{i=1}^k b_i e^{-2h/\phi_i}, \quad b_i = \frac{2\sigma_i^4}{\left(\sum_{j=1}^k \sigma_j^2\right)^2}.$$

### 4.5 Semilog transformation and scale identification

Writing  $y(h) = g(h) - 1$ , for two components we have

$$y(h) = b_S e^{-2h/\phi_S} + b_L e^{-2h/\phi_L}.$$

Taking logs,

$$\log y(h) = \log b_S - \frac{2h}{\phi_S} + \log\left(1 + \frac{b_L}{b_S} e^{-2h\left(\frac{1}{\phi_L} - \frac{1}{\phi_S}\right)}\right).$$

This is a log-sum-exp form: globally curved, but in limiting regimes it becomes linear:

$$\begin{aligned} \log y(h) &\approx \log b_S - \frac{2h}{\phi_S}, && \text{if small scale dominates,} \\ \log y(h) &\approx \log b_L - \frac{2h}{\phi_L}, && \text{if large scale dominates.} \end{aligned}$$

The crossover (“shoulder”) occurs at

$$h^* = \frac{\phi_S \phi_L}{2(\phi_L - \phi_S)} \log\left(\frac{b_S}{b_L}\right).$$

Thus, on a semilog plot of  $g(h) - 1$ , the CSCP naturally reveals its modular structure: straight-line regimes with slopes  $-2/\phi_S$  and  $-2/\phi_L$ , with the transition at  $h^*$ . See Figure 2.

**Reconstruction by two-line regression.** In Figure 2, the orange dotted line is an estimate of  $g(h) - 1$  obtained by a simple two-line procedure:

1. Select two ranges of  $h$ : one where the short-scale term dominates, another where the long-scale term dominates.
2. On each range, regress  $\log\{g(h) - 1\}$  on  $h$  to obtain a slope  $m$  and intercept  $c$ .
3. Translate these line parameters back into exponential form:

$$\hat{\phi} = -\frac{2}{m}, \quad \hat{b} = e^c.$$

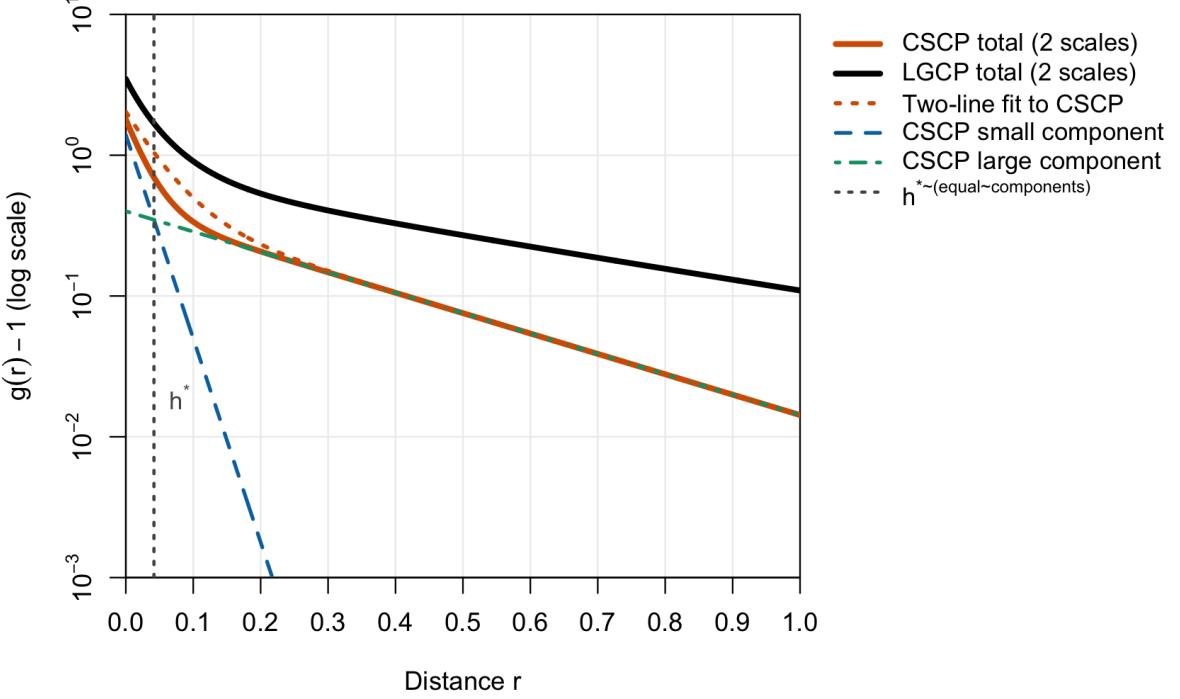


Figure 2: Semilogarithmic plot of  $g(h) - 1$  for a two-scale CSCP. Because  $g(h) - 1 = b_S e^{-2h/\phi_S} + b_L e^{-2h/\phi_L}$ , the function is a sum of exponentials. On the semilog scale, this produces two approximately linear regimes: one at small lags with slope  $-2/\phi_S$  (short-range clustering), and one at larger lags with slope  $-2/\phi_L$  (long-range clustering). The vertical marker denotes the crossover distance  $h^*$ , where the two component contributions are equal. This “shoulder” provides a natural diagnostic of the modular structure of CSCPs, in contrast to LGCPs where multiple scales are entangled in the exponent and the curve remains smoothly curved.

4. Combine the two estimated exponentials:

$$\hat{y}(h) = \hat{b}_S e^{-2h/\hat{\phi}_S} + \hat{b}_L e^{-2h/\hat{\phi}_L}.$$

This is capable of reproducing the true CSCP curve closely when the two scales are well separated. The key diagnostic feature is that the CSCP decomposes naturally into additive straight lines on the log scale, making its modular structure identifiable. In contrast, the LGCP curve (black) remains smoothly curved on this scale and does not admit such a decomposition.

## 5 Comparison: LGCP pair correlation

For a two-scale log-Gaussian Cox process (LGCP), the pair correlation is

$$g(h) = \exp\left\{C_S(h) + C_L(h)\right\},$$

where  $C_S(h), C_L(h)$  are the covariance functions of the short- and long-range Gaussian fields (e.g. exponential or Matérn).

**Log structure.** Unlike the CSCP, taking logarithms does not linearise  $g(h) - 1$ :

$$\log(g(h) - 1) \approx C_S(h) + C_L(h),$$

which is the *sum of covariance functions*, not a sum of exponentials. On a semilog plot of  $g(h) - 1$ , the curve remains smoothly curved — there are no clean straight-line regimes.

**No modular decomposition.** For the LGCP, the short- and long-scale contributions enter additively *inside the exponent*, so their effects are inseparably entangled. This produces a global smoothness: there is no finite interval of  $h$  where one scale contributes a straight line independent of the other.

**Contrast with CSCP.** The LGCP curve (black in Figure 2) bends continuously, while the CSCP curve (orange) is well approximated by the sum of two linear regimes on the semilog scale. This is the key diagnostic difference:

- **CSCP:** additive structure in  $g(h) - 1$  itself  $\rightarrow$  modular, identifiable scales.
- **LGCP:** additive structure in the exponent  $\rightarrow$  smooth entanglement of scales, no modular separation.

Thus, while the two models can look superficially similar on a linear scale, the semilog representation reveals that only the CSCP decomposes naturally into interpretable multi-scale contributions.

## 6 Dealing with the mean

We'd like to control the expected number of points, preferably decoupling from the variance(s). A natural first attempt at a stationary reparameterisation is to work with

$$\bar{\lambda}, \quad w \in [0, 1], \quad \{\alpha_i\}_{i=1}^k, \quad \sum_i \alpha_i = 1,$$

where  $\bar{\lambda}$  is the target mean intensity,  $w$  allocates the proportion of this mean explained by the stochastic CSCP component, and  $\alpha_i$  partitions  $w\bar{\lambda}$  across the  $k$  Gaussian-field summands. In this scheme one sets

$$\sigma_i^2 = w\bar{\lambda}\alpha_i, \quad \mu_i = 0,$$

so that the overall mean is respected.

**Limitations.** This parameterisation has several drawbacks:

- *Entanglement of mean and variance:* Since the variances  $\sigma_i^2$  are tied directly to  $\bar{\lambda}$ , increasing the expected number of points also changes the second-order clustering structure. Thus a dataset with 100 points and one with 10,000 points would imply different  $g(h)$  curves even if the underlying spatial correlation should be the same.
- *Lack of invariance:* For a Cox process one would ideally like the pair correlation  $g(h)$  to depend on correlation ranges  $\phi_i$  and relative weights  $\alpha_i$ , not on the overall sample size. The above construction does not preserve this invariance.
- *Identifiability issues:* Different combinations of  $(\bar{\lambda}, w, \alpha_i)$  can lead to the same mean intensity but distinct implied clustering, making fitted values hard to interpret. With large  $k$  this risks overfitting.

**Alternative: Mean-invariant scaling.** A cleaner approach is to decouple the first- and second-order structures. One may write

$$\Lambda(s) = \lambda_0 + \sum_{i=1}^k Z_i^2(s),$$

with  $\lambda_0$  as a separate baseline mean intensity, and  $Z_i$  Gaussian fields with variance parameters  $\sigma_i^2$  and ranges  $\phi_i$ . Here the expected total intensity is governed by  $\lambda_0 + \sum_i \sigma_i^2$ , while the pair correlation function is determined entirely by  $(\sigma_i^2, \phi_i)$ . This ensures that increasing  $\bar{\lambda}$  via  $\lambda_0$  does not alter the shape of  $g(h)$ , giving stable second-order behaviour across sample sizes.

**Caution on  $k$ .** Although the mean-invariant form has cleaner separation, identifiability remains delicate when many components are allowed. With  $k \geq 3$  it becomes increasingly difficult to distinguish overlapping scales, and parameter trade-offs between  $\lambda_0$  and  $\{\sigma_i^2\}$  can emerge. In practice, 2–3 components are likely the most defensible balance between flexibility and identifiability.

## 7 Fitting a multi-scale CSCP under mean-stationarity

For a stationary  $k$ -component CSCP, suppose we write the intensity as

$$\Lambda(s) = \lambda_0 + \sum_{i=1}^k Z_i(s)^2, \quad Z_i(s) \sim \text{GRF}(0, \sigma_i^2 \rho_i(\|h\|)),$$

where  $\lambda_0 \geq 0$  is a deterministic baseline, and the  $Z_i$  are independent zero-mean Gaussian random fields with marginal variances  $\sigma_i^2$  and correlation functions  $\rho_i(h)$ .

### 7.1 The mean–variance entanglement problem

If we set  $\lambda_0 = 0$ , then

$$\mathbb{E}\{\Lambda(s)\} = \sum_{i=1}^k \sigma_i^2,$$

and the pair correlation takes the form

$$g(h) - 1 = \frac{2 \sum_{i=1}^k \sigma_i^4 \rho_i(h)^2}{(\sum_{j=1}^k \sigma_j^2)^2}.$$

As noted above, the same parameters  $\sigma_i^2$  simultaneously control both the *first order* (mean intensity) and the *second order* (clustering strength). This is undesirable in practice: if the observed pattern has, say, 1000 points, we do not want all of that count to be explained entirely by the  $\sigma_i^2$ , since their role should be primarily to govern the second-order properties.

### 7.2 Trying to decouple the mean from the second order

Include a non-random baseline  $\lambda_0$ , and reparameterize as follows. Let  $\bar{\lambda}$  denote the desired stationary mean intensity of the process. Introduce:

- a mean fraction  $w \in (0, 1)$  specifying the “proportion of  $\bar{\lambda}$  explained by the random components”,
- weights  $\alpha_i \geq 0$  with  $\sum_{i=1}^k \alpha_i = 1$ , allocating that random fraction across components,
- a baseline  $\lambda_0 = (1 - w)\bar{\lambda}$ .

Then define

$$\sigma_i^2 = w \bar{\lambda} \alpha_i, \quad i = 1, \dots, k.$$

By construction,

$$\mathbb{E}\{\Lambda(s)\} = \lambda_0 + \sum_i \sigma_i^2 = (1 - w)\bar{\lambda} + w\bar{\lambda} = \bar{\lambda},$$

so the total mean intensity is decoupled and controlled directly by  $\bar{\lambda}$ . But there is still a clear link to  $\sigma_i^2$  (see above), so I’m not sure this will work.

### 7.3 Pair correlation under this parameterization

Substituting into the expression for  $g(h) - 1$ , we obtain

$$g(h) - 1 = 2w^2 \sum_{i=1}^k \alpha_i^2 \rho_i(h)^2.$$

This has the following features:

- The overall mean  $\bar{\lambda}$  no longer appears in the pair correlation, so estimation of second-order properties is stable and independent of the mean.
- The parameter  $w$  directly controls the overall clustering strength:  $w = 0$  gives a homogeneous Poisson process, while larger  $w$  increases the variance-to-mean ratio.
- The parameters  $\alpha_i$  and  $\phi_i$  (through  $\rho_i$ ) control the relative contributions and correlation scales of the components.

### 7.4 Practical implications

This reparameterization cleanly separates the roles of the parameters:

- $\bar{\lambda}$ : fixes the average number of points (first order).
- $w$ : governs the overall clustering amplitude.
- $\{\alpha_i, \phi_i\}$ : control the decomposition of clustering across scales.

Thus, the  $\sigma_i^2$  need not be “responsible” for the mean. Instead, they inherit values consistent with  $(\bar{\lambda}, w, \alpha_i)$ , leaving  $\phi_i$  and  $\alpha_i$  to shape the second-order structure. This addresses the entanglement problem and (maybe?) provides a principled, interpretable framework for fitting multi-scale CSCPs in practice.

### 7.5 Identifiability considerations

The proposed parameterisation separates the mean intensity  $\bar{\lambda}$ , the clustering amplitude  $w$ , and the component weights  $\alpha_i$ . In practice, their identifiability depends on which order of properties is being used:

- **First order:** The average point count identifies  $\bar{\lambda}$  directly, up to the window size. This is estimable from the data without reference to second-order structure.
- **Second order:** The pair correlation function depends only on the *relative* contributions of the random components:

$$g(h) - 1 = 2w^2 \sum_{i=1}^k \alpha_i^2 \rho_i(h)^2.$$

Thus, the  $w$  parameter is identifiable only through the *overall amplitude* of clustering, while the  $\alpha_i$  are identifiable up to their relative proportions given distinct correlation scales  $\phi_i$ .

- **Interaction of parameters:**  $\bar{\lambda}$  does not appear in  $g(h) - 1$ , so it is cleanly separated. However,  $w$  and the  $\alpha_i$  always enter together as the combination  $w^2 \alpha_i^2$ . Hence:
  - $w$  is estimable from the global variance-to-mean ratio (how clustered the process is overall),
  - $\{\alpha_i, \phi_i\}$  are estimable from the *shape* of  $g(h) - 1$  (multi-scale decay).

**Practical implication.** Identifiability is going to be bad in small samples: different  $(w, \alpha)$  combinations can mimic each other if the  $\phi_i$  are similar. But when scales are well separated, the log-representation of  $g(h) - 1$  should show clear linear regimes, making  $w$ ,  $\alpha_i$ , and  $\phi_i$  practically estimable. Thus, the reparameterisation is both interpretable and identifiable, provided that the components act on distinct spatial ranges.

## 7.6 Large $k$ is gonna be shithouse, almost surely

While the CSCP reparameterisation  $(\bar{\lambda}, w, \{\alpha_i, \phi_i\}_{i=1}^k)$  is interpretable, it's easy to see that identifiability will deteriorate as  $k$  grows. The reason is twofold:

1. The second-order signal is  $g(h) - 1 = 2w^2 \sum_{i=1}^k \alpha_i^2 \rho_i(h; \phi_i)^2$ . If several  $\phi_i$  are similar, their contributions are nearly collinear on  $\log\{g(h) - 1\}$ , making  $(\alpha_i, \phi_i)$  poorly separated. Finite window size, edge correction, and binning further reduce the ‘effective resolution’ of distinct decay rates.
2. Only the products  $w^2 \alpha_i^2$  (modulated by  $\rho_i^2$ ) enter  $g(h) - 1$ , so  $w$  and  $\{\alpha_i\}$  trade off unless the *shape* information (from distinct  $\phi_i$ ) is strong. With larger  $k$ , different mixtures can fit  $g(h)$  nearly equally well.

So, if using this version we should:

- **Prefer small  $k$  (e.g.  $k = 2$  or  $3$ ).** Choose the smallest  $k$  that explains the dominant slope regimes on the semilog plot of  $\hat{g}(h) - 1$ .
- **Enforce ordering and separation.** Constrain  $0 < \phi_1 < \dots < \phi_k$  and, if needed, a minimal separation (e.g.  $\phi_{i+1}/\phi_i \geq \delta > 1$ ) to avoid label switching and near-duplicates.
- **Penalise complexity.** Use information criteria tailored to second-order/composite fits (e.g. CL-AIC/CL-BIC) or cross-validated contrast on  $\hat{g}(h)$  when comparing  $k$ .
- **Stabilise weights.** Parameterise  $\alpha_i = \exp(\eta_i)/\sum_j \exp(\eta_j)$  and add a mild ridge penalty on  $\eta$  (or a Dirichlet prior in a Bayesian fit) to discourage spurious tiny components.
- **Use robust initialisation.** Detect slope plateaus on  $\log\{\hat{g}(h) - 1\}$  via sliding-window regressions and change-point segmentation; initialise  $(\phi_i, \alpha_i)$  from these plateaus, then refine.
- **Quantify uncertainty.** Report bootstrap intervals for  $(\phi_i, \alpha_i, w)$  or standard errors from a pairwise/composite likelihood. Wide intervals or unstable plateaus will be a red flag for excessive  $k$ .

**Rule of thumb:** We could recommend adopting  $k = 2$  when a clear “steep then shallow” pattern appears; consider  $k = 3$  only if a third, visibly distinct linear regime is sustained over a nontrivial range of  $h$ . Otherwise, additional components are likely unidentifiable and amount to overfitting noise.

## 7.7 Dealing with $w = 1$

In the current parameterisation, the random part of the intensity contributes a fraction  $w \in [0, 1]$  of the total mean, with the remaining fraction  $(1 - w)$  assigned to a homogeneous baseline. The special case  $w = 1$  corresponds to a *fully stochastic CSCP* in which *all points* arise from the Gaussian random fields, with no superimposed baseline Poisson component.

**Definition.** Let  $\bar{\lambda}$  denote the desired stationary mean intensity, and let  $\alpha_1, \dots, \alpha_k$  be non-negative weights summing to one. Then define

$$\sigma_i^2 = \bar{\lambda} \alpha_i, \quad \mu_i = 0, \quad Z_i(s) \sim \text{GRF}(0, \sigma_i^2 \rho_i(\|h\|)).$$

The intensity field is

$$\Lambda(s) = \sum_{i=1}^k Z_i(s)^2.$$

**First-order behaviour.** By construction,

$$\mathbb{E}\{\Lambda(s)\} = \sum_{i=1}^k \sigma_i^2 = \bar{\lambda},$$

so the expected number of points in a region  $W$  is  $\bar{\lambda}|W|$ .

**Second-order behaviour.** The pair correlation function is

$$g(h) - 1 = \frac{2 \sum_{i=1}^k \sigma_i^4 \rho_i(h)^2}{\left( \sum_{j=1}^k \sigma_j^2 \right)^2} = 2 \sum_{i=1}^k \alpha_i^2 \rho_i(h)^2.$$

Crucially, the factor  $\bar{\lambda}$  cancels: *the shape and strength of clustering depend only on the weights  $\alpha_i$  and correlation functions  $\rho_i$ , not on the overall mean intensity.* Thus the pair correlation is invariant to changes in  $\bar{\lambda}$ .

**Implications for the number of points.** The expected count scales linearly with  $\bar{\lambda}$ , but the form of  $g(h)$  is unchanged. However, as in any Cox process, the variance of the total count grows faster:

$$\text{Var}(N) = \bar{\lambda}|W| + \bar{\lambda}^2 C_W,$$

where  $C_W = \iint_W c(\|s-t\|) ds dt$  depends on the kernel  $c(h) = 2 \sum_i \alpha_i^2 \rho_i(h)^2$ . Consequently, the index of dispersion,  $\text{Var}(N)/\mathbb{E}[N]$ , increases linearly with  $\bar{\lambda}$ . This reflects the natural property that larger populations exhibit more realisation-to-realisation variability, even though the spatial clustering structure is unchanged.

**Interpretation.** The case  $w = 1$  is especially appealing when the scientific assumption is that *all observed points arise from clustered structure*, without a homogeneous background. The parameters then separate cleanly as follows:

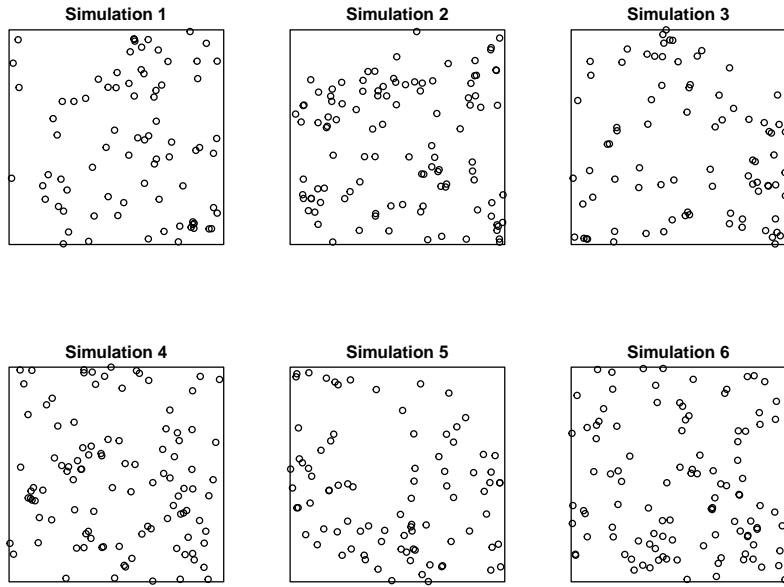
- $\bar{\lambda}$  controls the expected number of points,
- $\alpha_i$  and  $\rho_i$  control the relative contribution and spatial range of clustering at each scale.

Figure 4 gives realisations of two 2-component CSCPs, one set with  $\bar{\lambda} = 100$ , the other with  $\bar{\lambda} = 10000$ .

## 7.8 Dealing with $0 < w < 1$

Reminder:  $w \in [0, 1]$  specifies the fraction of the mean intensity carried by the *clustered* (random) part, with the remaining  $(1-w)$  assigned to a homogeneous baseline.

reparam CSCP:  $\bar{\lambda}=100$ ;  $w=1$ ;  $\alpha=0.5,0.5$ ;  $\phi=0.02,0.2$



reparam CSCP:  $\bar{\lambda}=10000$ ;  $w=1$ ;  $\alpha=0.5,0.5$ ;  $\phi=0.02,0.2$

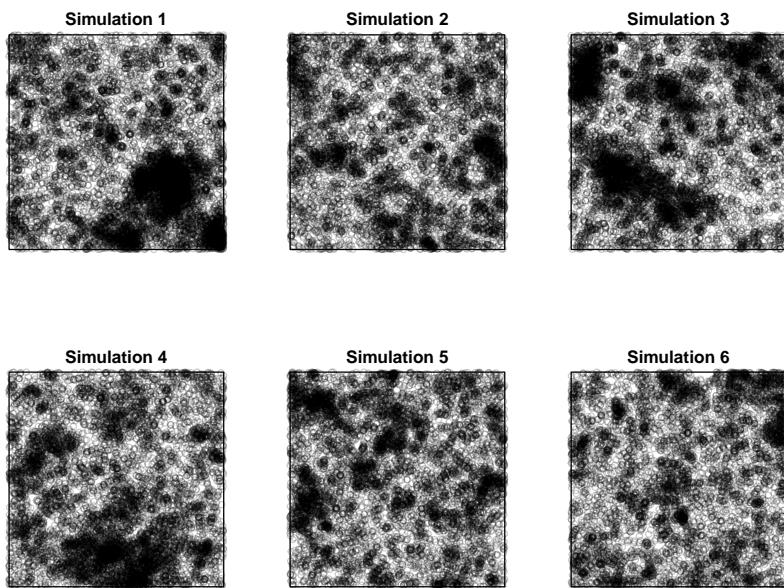


Figure 3: Two sets of realisations (six each) for  $\bar{\lambda} \in \{100, 10000\}$ , of a reparameterised  $k = 2$  CSCP with  $w = 1$  (fully stochastic).

**Definition.** Fix a target mean intensity  $\bar{\lambda} > 0$  and weights  $\alpha_1, \dots, \alpha_k \geq 0$  with  $\sum_i \alpha_i = 1$ . Let

$$\sigma_i^2 = w \bar{\lambda} \alpha_i, \quad \mu_i = 0, \quad Z_i(s) \sim \text{GRF}(0, \sigma_i^2 \rho_i(\|h\|)),$$

and set the baseline to

$$\lambda_0 = (1 - w) \bar{\lambda}.$$

The intensity field is

$$\Lambda(s) = \lambda_0 + \sum_{i=1}^k Z_i(s)^2.$$

**First-order behaviour.** By construction,

$$\mathbb{E}\{\Lambda(s)\} = \lambda_0 + \sum_i \sigma_i^2 = (1 - w) \bar{\lambda} + w \bar{\lambda} = \bar{\lambda},$$

so the expected number of points in a window  $W$  is  $\bar{\lambda}|W|$ .

**Second-order behaviour (pair correlation).** For a Cox process,  $g(h) - 1 = \text{Cov}\{\Lambda(s), \Lambda(t)\}/\mathbb{E}\Lambda^2$ . Since the baseline is nonrandom, only the squared-field sum contributes to the covariance. Using independence of the  $Z_i$  and Isserlis' theorem,

$$g(h) - 1 = 2w^2 \sum_{i=1}^k \alpha_i^2 \rho_i(h)^2.$$

Hence:

- the *shape* across distances is governed by the scales  $\rho_i(\cdot)$  and the composition  $\{\alpha_i\}$ ;
- the *amplitude* of clustering is tuned by  $w$  (quadratically), with  $w \downarrow 0$  driving  $g(h) \rightarrow 1$  (homogeneous Poisson).

**Implications for the number of points.** The mean count remains  $\mathbb{E}[N] = \bar{\lambda}|W|$ , independently of  $w$ . The count variance decomposes as

$$\text{Var}(N) = \bar{\lambda}|W| + (w \bar{\lambda})^2 C_W, \quad C_W = \iint_W 2 \sum_{i=1}^k \alpha_i^2 \rho_i(\|s-t\|)^2 ds dt,$$

so the index of dispersion is

$$\frac{\text{Var}(N)}{\mathbb{E}[N]} = 1 + w^2 \bar{\lambda} \frac{C_W}{|W|}.$$

Thus, decreasing  $w$  damps both small-lag clustering (via  $g - 1$ ) and the between-realisation overdispersion of the total count, while leaving the mean unchanged.

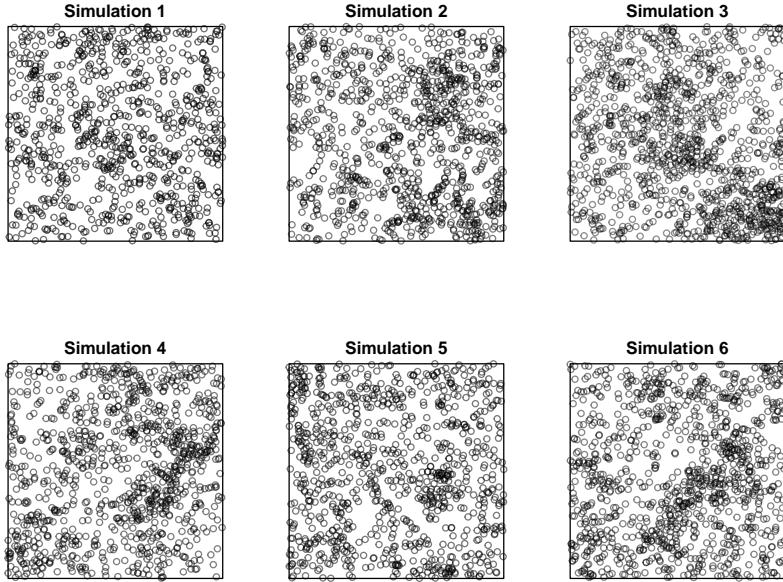
**Interpretation and use.** The baseline-plus-clustered form is useful when scientific knowledge suggests a nonzero background of “uninformative” events *in addition* to clustered mechanisms. Parameters separate cleanly:

- $\bar{\lambda}$  sets the expected number of points (first order);
- $w$  controls *how clustered* the pattern is overall (amplitude);
- $\{\alpha_i, \rho_i\}$  determine *where and at what scales* clustering occurs (shape).

As  $w \rightarrow 1$  we recover the fully stochastic CSCP (no baseline); as  $w \rightarrow 0$  the model reduces to a homogeneous Poisson process.

Figure ?? again gives realisations of two 2-component CSCPs, this time with one at  $w = 0.75$ , the other with  $w = 0.1$  (effectively homogeneous Poisson).

reparam CSCP:  $\bar{\lambda}=1000$ ;  $w=0.75$ ;  $\alpha=0.5,0.5$ ;  $\phi=0.02,0.2$



reparam CSCP:  $\bar{\lambda}=1000$ ;  $w=0.1$ ;  $\alpha=0.5,0.5$ ;  $\phi=0.02,0.2$

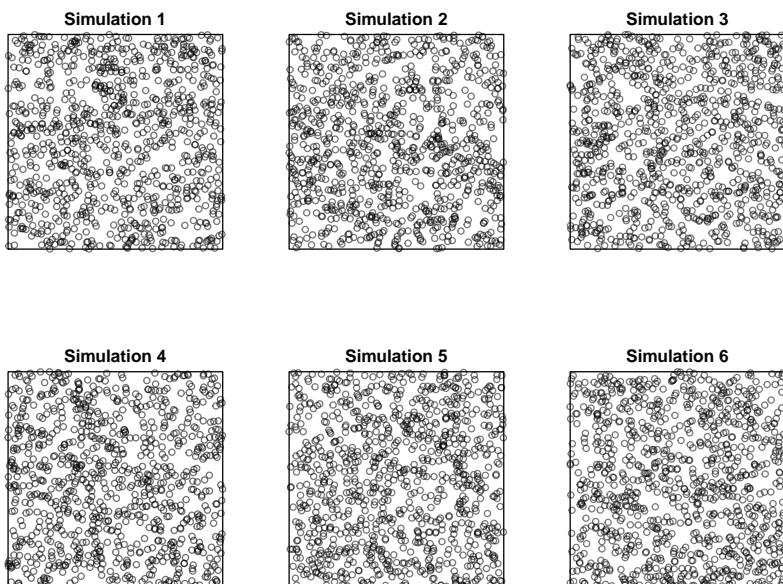


Figure 4: Two sets of realisations (six each) for  $w \in \{0.75, 0.1\}$ , of a reparameterised  $k = 2$  CSCP.

## 8 Fitting Strategy for CSCPs ( $w = 1$ )

We consider the stationary chi-squared Cox process

$$\Lambda(s) = \sum_{i=1}^k Z_i(s)^2, \quad \text{Cov}\{Z_i(s), Z_i(s+h)\} = \sigma_i^2 \rho_i(h; \phi_i),$$

where the  $\{Z_i\}$  are independent, mean-zero Gaussian random fields. Define

$$\alpha_i = \frac{\sigma_i^2}{\sum_j \sigma_j^2}, \quad \sum_i \alpha_i = 1, \quad \bar{\lambda} = \mathbb{E}[\Lambda] = \sum_i \sigma_i^2 \approx \frac{n}{|W|}.$$

With  $w = 1$ , the pair correlation simplifies to

$$g(h) - 1 = 2 \sum_{i=1}^k \alpha_i^2 \rho(h; \phi_i)^2,$$

so second-order structure is controlled entirely by  $(\alpha, \phi)$ , while  $\bar{\lambda}$  fixes the overall scale through  $\sigma_i^2 = \bar{\lambda} \alpha_i$ .

### 8.1 Case $k = 1$ : Single-Component CSCP

When  $k = 1$ , the intensity is  $\Lambda(s) = Z(s)^2$  and

$$g(h) - 1 = 2 \rho(h; \phi)^2.$$

For an exponential correlation,  $\rho(h) = \exp(-h/\phi)$ , this reduces to

$$g(h) - 1 = 2 \exp(-2h/\phi), \quad \log\{g(h) - 1\} = \log 2 - \frac{2}{\phi}h,$$

a perfect straight line on semilogarithmic axes.

**Fitting procedure ( $k = 1$ ):**

1. Compute empirical  $\hat{g}(h)$ , set  $y(h) = \max\{\hat{g}(h) - 1, \varepsilon\}$ .
2. Fit a linear regression  $\log y(h) = c + mh$  over a stable mid-range band  $[a, b]$ .
3. Estimate  $\hat{\phi} = -2/m$ . The intercept should be close to  $\log 2$ .
4. Set  $\hat{\sigma}^2 = \hat{\lambda}$ .
5. Validate using bootstrap envelopes and by checking linearity of  $\log\{\hat{g}(h) - 1\}$ .

### 8.2 Case $k = 2$ or $3$ : Multi-Component CSCP

For  $k = 2$ ,

$$g(h) - 1 = 2\alpha_1^2 \rho(h; \phi_1)^2 + 2\alpha_2^2 \rho(h; \phi_2)^2,$$

and similarly for  $k = 3$ . The log–pcf is a *log-sum-exp* of exponentials, producing “shoulders” in semilog plots where small-scale and large-scale components trade dominance.

**Fitting procedure ( $k = 2, 3$ ):**

1. Compute  $\hat{g}(h)$  and set  $y(h) = \max\{\hat{g}(h) - 1, \varepsilon\}$ .
2. On semilog scale, choose disjoint distance bands (small, medium, large).
3. Fit straight lines  $\log y(h) \approx c_i + m_i h$  in each band.
4. Initialise  $\phi_i^{(0)} = -2/m_i$  and  $\alpha_i^{(0)} \propto \sqrt{e^{c_i}/2}$  (normalised).
5. Refine with constrained non-linear least squares (or minimum contrast) enforcing  $\alpha_i \geq 0$ ,  $\sum \alpha_i = 1$ , and ordered scales  $\phi_1 < \phi_2 (< \phi_3)$ .
6. Map back:  $\sigma_i^2 = \bar{\lambda} \alpha_i$ .

### 8.3 Uncertainty and Diagnostics

- **Uncertainty:** Use block or toroidal bootstrap, or pairwise composite likelihood with Godambe information.
- **Diagnostics:**
  - For  $k = 1$ ,  $\log\{g(h) - 1\}$  should be nearly linear.
  - For  $k = 2, 3$ , look for piecewise linear regimes on the semilog plot.
  - Simulated envelopes for  $g$ ,  $K$ , and  $J$  provide further checks.

### 8.4 Practical Notes

- $k = 1$  serves as a natural baseline; higher  $k$  allows modular separation of scales.
- Identifiability improves if  $\phi_i$  are well separated.
- Small  $k$  (2–3) is defensible; larger  $k$  risks overfitting and instability.

**Intuition: Why semilog regression works for CSCPs.**

- *Case  $k = 1$ :* The pair correlation reduces to

$$g(h) - 1 = 2e^{-2h/\phi}, \quad (1)$$

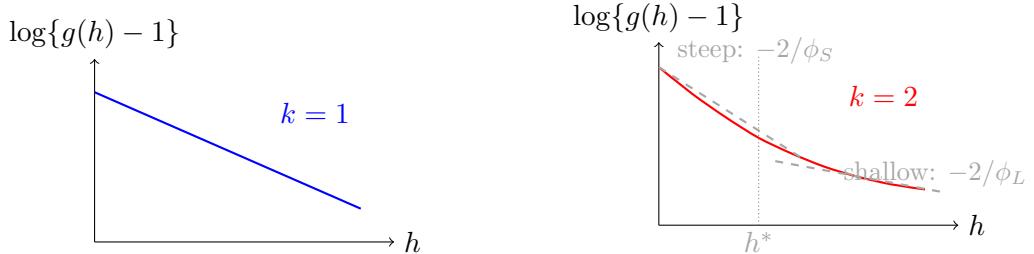
which is exactly an exponential decay. On semilog axes,  $\log\{g(h) - 1\}$  is a straight line with slope  $-2/\phi$ . *Interpretation:* A single scale of clustering produces a single straight line.

- *Case  $k = 2$ :* Now

$$g(h) - 1 = 2\alpha_1^2 e^{-2h/\phi_1} + 2\alpha_2^2 e^{-2h/\phi_2},$$

a *sum of two exponentials*. On semilog axes this becomes a log-sum-exp curve: linear at small  $h$  (dominated by the short range), linear again at large  $h$  (dominated by the long range), with a visible “shoulder” where the two contributions cross over. *Interpretation:* Two scales of clustering appear as two approximately straight lines joined by a bend.

**Schematic illustration (with guide-lines):**



**Practical implication:** Semilog regression turns multi-scale detection into searching for *piecewise linear* regimes. A straight line  $\Rightarrow$  single scale; a bend with two dashed guide-lines  $\Rightarrow$  two scales (short- then long-range dominance).

## 9 Fitting the single-component CSCP ( $k = 1$ )

Let's focus on the simplest case of  $k = 1$ , as detailed a moment ago in Section 8.1. Reminder: the theoretical pair correlation function of the CSCP with  $w = 1$  can be written as

$$g(r) - 1 = b \exp\left(-\frac{2r}{\phi}\right),$$

where  $\phi$  is the correlation range parameter, and  $b$  is a second-order scaling constant determined by the reparameterised model. Taking logarithms yields a linear form:

$$\log\{g(r) - 1\} = \log b - \frac{2}{\phi} r.$$

This motivates a simple semilog regression estimator: fit a straight line to the empirical curve  $\log(\hat{g}(r) - 1)$  as a function of  $r$ .

### 9.1 Interpretation

Suppose the regression gives slope  $\hat{\beta}_1$  and intercept  $\hat{\beta}_0$ . Then:

$$\hat{\phi} = -\frac{2}{\hat{\beta}_1}, \quad \hat{b} = \exp(\hat{\beta}_0).$$

In the reparameterised stationary CSCP, the bump height  $b$  is not a free parameter. Instead, with target mean intensity  $\bar{\lambda}$ , stochastic fraction  $w \in [0, 1]$ , and component weight  $\alpha_1$  (with  $\alpha_1 = 1$  when  $k = 1$ ), the variance of the underlying Gaussian field is  $\sigma^2 = w\bar{\lambda}\alpha_1$ , leading to

$$b = \frac{2\sigma^4}{\bar{\lambda}^2} = 2w^2\alpha_1^2.$$

Thus, in the  $w = 1, k = 1$  case, the intercept estimator provides a direct method to recover the component weight:

$$\hat{\alpha}_1 = \sqrt{\frac{\hat{b}}{2}}. \tag{2}$$

### 9.2 Quick example

Figure 5 shows a realisation of a  $w = 1, k = 1$  CSCP, along with a plot of the empirical pcf (scaled as  $\log\{\hat{g} - 1\}$ ). Two simple linear models are fitted via ordinary least squares – one in which the intercept is locked at  $\log 2$  (see the upcoming simulations), the other where the intercept is allowed to be free. The estimate of  $\phi$  (and, in the case of a free intercept,  $\alpha_1$ ) are obtained deterministically from these regressions.

### 9.3 Simulation 1: free intercept

I implemented the above strategy in a simulation study, generating replicated CSCP patterns in the unit square under known  $\phi = 0.1$ ,  $\bar{\lambda} = 1000$  (with  $k = 1$  we also have  $w = 1$  and  $\alpha_1 = 1$ ) and then applied the regression approach to each replicate. I used the divide-by-A estimator of the pcf, with Jones-Foster boundary correction and Ripley edge correction, combined with the least-squares estimate of the pcf bandwidth; then,  $\hat{g}$  replaces  $g$  in the regression equation. Figure 6 (left) shows the histogram of estimated  $\phi$ , which is centred close to the true value with a mild right skew. This behaviour is expected: slope information is drawn from the tail of the empirical semilog plot, which is relatively stable even under finite sample variability. Thus,  $\phi$  is recovered reliably.

Figure 6 (right) shows the corresponding histogram of the derived  $\hat{\alpha}_1$ . In theory  $\alpha = 1$ , yet in practice we observed some variability. The value of  $\hat{\alpha}_1$  should not be treated as a free

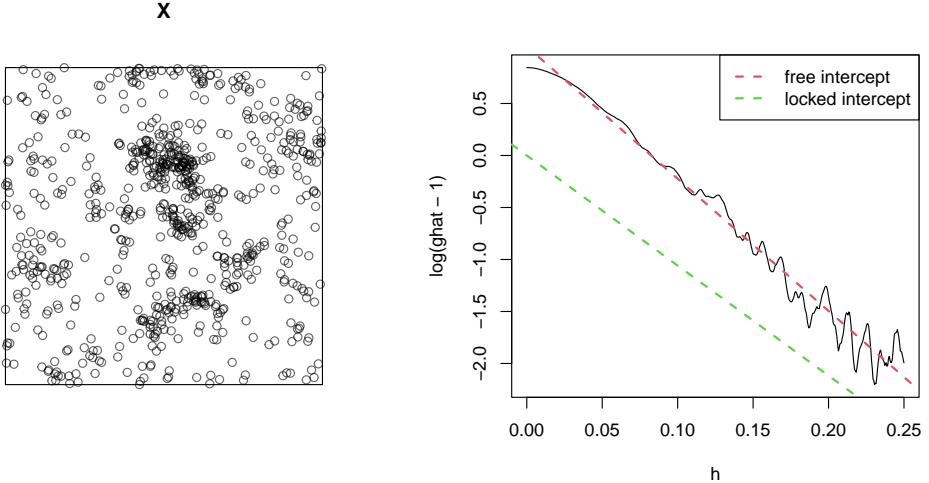


Figure 5: Example CSCP realisation with corresponding semilogarithmic empirical pcf, to which simple linear regressions are applied.

parameter in the  $k = 1$  case, but rather as a diagnostic check. Ideally, the histogram of  $\hat{\alpha}$  should be centred near one, confirming that the regression procedure is internally consistent. In this case, we see satisfactory behaviour.

Simple linear regression on the semilog scale could provide a pretty good method for recovering the correlation range  $\phi$  in the single-component case. The corresponding derivations of the  $\alpha_1$  term are less reliable – in part these will reflect shortcomings of the estimator of the pcf itself – but can be used to check for consistency, since in the  $k = 1$  model they should clearly cover unity.

#### 9.4 Simulation 2: locked intercept

Given it's determined by the model, why estimate the intercept of the regression on  $\log\{\hat{g}(r) - 1\}$  at all? Under the current definitions,  $b \equiv 2$  as per (1), and so the intercept has to be  $\log 2 \approx 0.693$ ; indeed, we can see this in (2), which leads to  $\alpha_1 \equiv 1$ , matching the model definition.

The current set of simulations is repeated on the same set of simulated datasets, except now, the intercept is held at  $\log 2$  as a fixed offset, and we fit the line searching only for the optimal slope  $\hat{\beta}_1$ . The hope is that this will lead to more precise estimation of  $\phi$ . Will deficiencies in estimation of the pcf harm things?

Figure 7 gives the resulting estimates of  $\phi$ , again these look good – centered on the truth. The RMSE is also notably smaller, suggesting the locked intercept is helpful.

## 10 Positioning relative to existing work

The model we describe as a *Chi-squared Cox process (CSCP)* has clear links to the *permanental process* introduced by McCullagh and Møller (2006) [2], and previously foreshadowed in the boson process literature [3]. Their work established the existence, joint densities, and moment properties of processes with intensity fields of the form  $\Lambda(s) = \sum_{i=1}^k Z_i(s)^2$ , with  $Z_i$  Gaussian random fields. In this sense, the theoretical foundation of CSCPs is already solidly in place.

What seems to be missing, however, is applied development. In contrast to log-Gaussian Cox processes (LGCPs), permanental/CSCPs have seen almost no uptake in spatial statistics practice. I think. In particular, several potentially valuable aspects have not been explored:

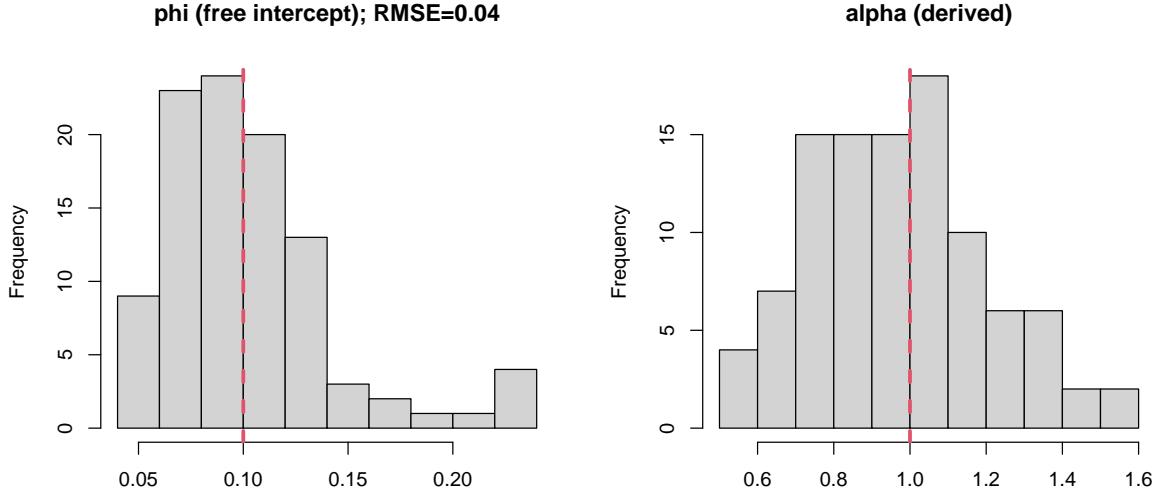


Figure 6: Simulation study with  $k = 1$ . Left: estimates of  $\phi$  from regression of  $\log\{\hat{g}(r) - 1\}$  on  $r$ . Right: derived estimates of  $\alpha$  from regression intercepts. Dashed lines mark true/target values.

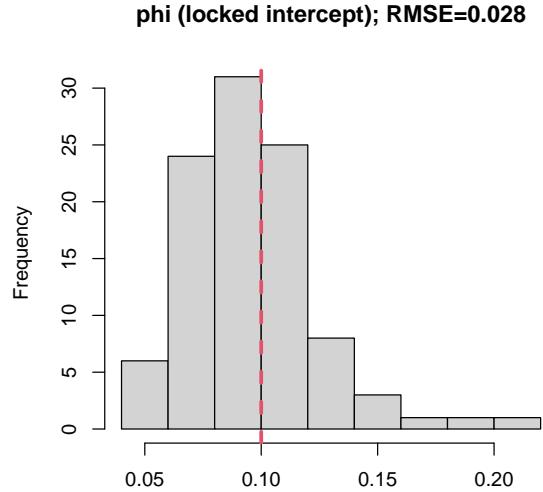


Figure 7: Simulation study with  $k = 1$ . Left: estimates of  $\phi$  from regression of  $\log\{\hat{g}(r) - 1\}$  on  $r$ , with the intercept held fixed at  $\log 2 \implies \alpha_1 = 1$ .

- **Diagnostics and interpretability:** For exponential correlations,  $g(h) - 1$  decomposes as a sum of exponentials, which on a semilog plot reveals straight-line regimes with slopes  $-2/\phi_i$ . This provides a natural diagnostic and a modular interpretation of multiple scales of clustering — a feature absent in LGCPs, where scales combine inside the exponent.
- **Multi-scale modelling:** The modular structure suggests CSCP may offer advantages in modelling both small-scale hotspots and large-scale background variation simultaneously. Identifying these scales empirically and contrasting them with LGCP fits has not yet been demonstrated.

- **Estimation and heuristics:** Practical inference strategies (e.g. composite likelihood, minimum contrast, or heuristic initialisation from pair correlation diagnostics) remain underexplored.
- **Applied case studies:** To our knowledge, there are no published applications of CSCPs to real epidemiological or ecological point pattern datasets.

Accordingly, while the permanental/CSCP model itself is not new (some other relevant papers are [4, 5, 6]), there is possibly scope for novel contributions in terms of applied methodology, diagnostics, and interpretability. Our work is positioned in this gap: to revisit CSCPs with emphasis on their practical merits and to provide the first systematic comparison with LGCPs in applied settings.

## References

- [1] J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, 2003. Available at Routledge: <https://doi.org/10.1201/9780203496930>.
- [2] P. McCullagh and J. Møller. The permanental process. *Advances in Applied Probability*, 38(4):873–888, 2006. Open version: Cambridge Core PDF <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/1E60BA9879EAE9BBF84321FA390AFD81/S0001867800001361a.pdf/the-permanental-process.pdf>.
- [3] T. Shirai and Y. Takahashi. Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *Journal of Functional Analysis*, 205(2):414–463, 2003. Publisher page: <https://www.sciencedirect.com/science/article/pii/S002212360300171X>.
- [4] C. J. Walder and A. N. Bishop. Fast Bayesian Intensity Estimation for the Permanental Process. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 2017. Open PDF: <https://proceedings.mlr.press/v70/walder17a/walder17a.pdf>.
- [5] O. Nicolis, L. Fernández, and J. E. Muñoz. Temporal Cox Process with Folded Normal Intensity. *Axioms*, 11(10):513, 2022. Open access: <https://www.mdpi.com/2075-1680/11/10/513>.
- [6] N. Eisenbaum. A Cox Process Involved in the Bose–Einstein Condensation. *Annales Henri Poincaré*, 9(6):1123–1140, 2008. Open PDF: <https://link.springer.com/content/pdf/10.1007/s00023-008-0376-6.pdf>.