# Developments in statistical inference when assessing spatiotemporal disease clustering with the tau statistic

Timothy M. Pollington [a,c,*], Michael J. Tildesley [b],
T. Déirdre Hollingsworth [c,1], Lloyd A.C. Chapman [d,1]

[a] *MathSys CDT, University of Warwick, UK*
[b] *Zeeman Institute (SBIDER), School of Life Sciences, and Mathematics Institute, University of Warwick, UK*
[c] *Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, UK*
[d] *London School of Hygiene & Tropical Medicine, UK*

## ARTICLE INFO

## ABSTRACT

The *tau statistic* $\tau$ uses geolocation and, usually, symptom onset time to assess global spatiotemporal clustering from epidemiological data. We test different methods that could bias the clustering range estimate based on the statistic or affect its apparent precision, by comparison with a baseline analysis of an open access measles dataset.

From re-analysing this data we find evidence against no clustering and no inhibition, *p*-value $\in [0, 0.022]$ (global envelope test). We develop a tau-specific modification of the Loh & Stein spatial bootstrap sampling method, which gives bootstrap tau estimates with 24% lower sampling error and a 110% higher estimated clustering endpoint than previously published (61·0 m *vs.* 29 m) and an equivalent increase in the clustering area of elevated disease odds by 342%. These differences could have important consequences for control efforts.

Correct practice of graphical hypothesis testing of no clustering and clustering range estimation of the tau statistic are illustrated in the online Graphical abstract. We advocate proper implementation of this useful statistic, ultimately to reduce inaccuracies in control policy decisions made during disease clustering analysis.

---

* Corresponding author at: MathSys CDT, University of Warwick, UK.
 *E-mail address:* timothy.pollington@gmail.com (T.M. Pollington).
1 Equal contributions from TDH & LACC.

## 1. Introduction

Assessing if *spatiotemporal clustering* is present and measuring its magnitude and range is informative for epidemiologists working to control infectious diseases. The *tau statistic* (Section 2) is more appropriate than most statistics for this task as it measures spatiotemporal rather than just spatial clustering, produces non-parametric estimates (without process assumptions) and, unlike the *K* function (Gabriel and Diggle, 2009), offers a relative magnitude in the difference of risk, rate or odds of disease versus the background level (Section 2.1) (Lessler et al., 2016; Pollington et al., 2019a). The tau statistic herein should not be confused with 'Kendall's tau statistic/rank correlation coefficient' (Bland, 2000). This study is motivated by a review of its use that found that its current implementation inflates type I errors (incorrectly rejecting a true null hypothesis) when testing for no clustering and no inhibition, and may bias estimates of the range of clustering (Pollington et al., 2019a).

We investigate these aspects by analysing a well-studied open access measles dataset containing household geolocations and symptom onset times of cases (Section 3.1). It represents a *spatially discrete process* since infection is only recorded and can only occur at discrete household locations, so the (statistical) support is not spatially continuous (Diggle et al., 2010).

We adopt an ordered approach: we first test for no clustering and no inhibition (Section 3.3) and then, conditional on finding evidence against this null hypothesis, we estimate the clustering endpoint and its first sampling error estimate (Section 3.4; online Graphical abstract). This approach is contrary to the current methods applied to the tau statistic and similar statistics (Pollington et al., 2019a), which incorrectly combine graphical hypothesis testing for no clustering and estimation of the clustering range (Section 3.2). We hope these improved methods will encourage proper application of this burgeoning statistic.

## 2. The tau statistic

The tau statistic $\tau$ is a non-parametric global clustering statistic which takes a disease frequency measure (risk, odds or rate) within a certain annulus around a case and compares it to the background measure (at any distance) and averaged over all cases (Salje et al., 2012; Lessler et al., 2016; Pollington et al., 2019a). It measures the tendency of case pairs to spatially cluster while implicitly accounting for how related they are in terms of transmission using temporal information, making it a *spatiotemporal* statistic.

### 2.1. Tau statistic (odds ratio estimator)

We describe the most common tau estimator $\hat{\tau}_{\text{odds}}$, which is based on the relative odds of disease (Lessler et al., 2016), rather than other forms of the statistic (including a new rate ratio estimator), which are described in a detailed review (Pollington et al., 2019a) from which this subsection draws heavily.

The distance form of the tau statistic $\tau_{\text{odds}}$ is the ratio of (i) the odds $\theta(d_l, d_m)$ of finding any case $j$ that is 'related' to any case $i$, within a half-closed annulus $[d_l, d_m)$, $(l, m \in \mathbb{Z}^+, m = l + 1)$, around case $i$, to (ii) the odds $\theta(0, \infty)$ of finding any case $j$ related to any case $i$ at any distance separation $(d_{ij} \geq 0)$ for $n$ total cases (Eq. (1) & Fig. 1).

$$
\hat{\tau}_{\text{odds}}(d_l, d_m) := \frac{\hat{\theta}(d_l, d_m)}{\hat{\theta}(0, \infty)}
$$

$$
\text{where } \hat{\theta}(d_l, d_m) = \frac{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \mathbb{1}(z_{ij} = 1, d_l \leq d_{ij} < d_m)}{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \mathbb{1}(z_{ij} = 0, d_l \leq d_{ij} < d_m)}
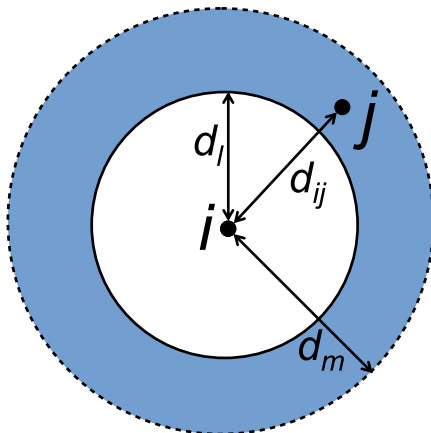$$

(1)

**Fig. 1.** A single distance band $[d_l, d_m]$, the primary argument of the distance form of a tau estimator function: a half-closed annulus of radii $d_l, d_m$ with a case $j$ inside, around case $i$, separated by distance $d_{ij}$.

The odds estimate $\hat{\theta}$ in Eq. (1) is the ratio of the number of related case pairs ($z_{ij} = 1$) within $[d_l, d_m]$, versus the number of unrelated case pairs ($z_{ij} = 0$) within $[d_l, d_m]$. The main computation is effectively a double sum over 'relatedness' indicator functions $\mathbb{1}(\cdot)$ for case pairs.

$\hat{\tau}(d_l, d_m)$ is then evaluated over $B$ total distance bands to give a *distance band set* $\underline{\triangle} :=$ $\{[d_l, d_m], l, m \in \mathbb{Z}^+, m = l + 1, l \leq B\}$. Tau values signify either the presence of spatiotemporal clustering ($\tau > 1$), no clustering and no inhibition ($\tau = 1$) or inhibition ($\tau < 1$). Sometimes an expanding disc is described by setting $d_l = 0$ and relabelling $d_m = d$ to give $\hat{\tau}(d)$ instead. Although $\hat{\tau}$ is strictly evaluated for a given distance band $[d_l, d_m]$, when a $\tau$-distance graph is drawn a value of $\hat{\tau}(d)$ can be obtained through linearly interpolating between the distance band endpoints. The half-closed annulus is a correction to the original open interval (Lessler et al. (2016): appendix 5); it was incorporated in December 2018 into the IDSpatialStats *R* package (which calculates the tau statistic) (Giles et al., 2018, 2019).

The relatedness of a case pair $z_{ij}$ is commonly determined using temporal information (*e.g.* difference in onset times of cases $i$ and $j$, *i.e.* $t_j - t_i$) (Pollington et al., 2019a). The *serial interval* is the period between the onset times of symptoms in the infector $t_i$ and their infectee $t_j$. Typically cases are defined as being temporally related when their onset times are within a single mean serial interval of each other.

$$ * \tau * $$

In the following sections (Sections 3–4) we provide a descriptive analysis of the data, before systematically testing several aspects of the tau statistic's implementation and their impact on the estimated clustering range and its sampling error.

## 3. Methods

### 3.1. The dataset and baseline analysis

We analyse an infectious disease dataset of measles from case households in Hagelloch, Germany in 1861 (Pfeilsticker, 1863; Oesterle, 1992; Neal and Roberts, 2004; Höhle et al., 2019). Computations were coded in *R* (R Core Team, 2019) with further detail in the Supplementary material. We have reproduced Lessler et al.'s (unpublished) analysis as a baseline result (Fig. 3); using *their* interpretation of Fig. 3, spatiotemporal clustering is reported up to $\sim 30$ m (Lessler et al., 2016). As $\tau$ is a global statistic, ideally we would explore a distance band set covering the majority of pairwise distances ($\sim 200$ m). However here we restrict the plots to 120 m for diagnostic (Fig. 3) or point estimation (Figs. 5, 6, 8 & B.3) purposes, to be consistent with the baseline analysis.

*3.2. Our approach to graphical hypothesis testing and point estimation*

An *envelope* is loosely defined as a collection of connected-line (*syn.* piecewise linear) functions in the Cartesian plane, with some bound applied above and below. *Central/null envelopes* describe the line function, *i.e.* whether it originates from bootstrap simulations of a point estimate or *time-mark permuted* null distribution (Section 3.3), respectively; whereas *global envelope* or *pointwise confidence interval* (*syn.* confidence band) refer to the way line functions are bounded. A global envelope is a confidence interval (CI) for a collection of line functions but does not represent a single distance band of one tau point estimate $\hat{\tau}(d_l, d_m)$ (*i.e.* a pointwise CI), but rather the entire distance band set $\underline{\Delta}$. At say a 95% significance level, in 95% of outcomes of constructing a global envelope, the random envelope would contain the true value of $\tau(d_l, d_m)$, $\forall [d_l, d_m] \in \underline{\Delta}$ (Baddeley et al., 2015).

Our graphical hypothesis test (Section 3.3) and point estimation methods (Section 3.4; online G raphical abstract) offer corrections to the implementations of the tau statistic or similar statistics used in many papers reviewed in Pollington et al. (2019a), *i.e.* (Salje et al., 2012, 2016a,b, 2017, 2018; Bhoomiboonchoo et al., 2014; Grabowski et al., 2014; Levy et al., 2015; Grantz et al., 2016; Hoang Quoc et al., 2016; Lessler et al., 2016; Azman et al., 2018; Rehman et al., 2018; Succo et al., 2018; Truelove et al., 2019). These studies incorrectly used an envelope about the point estimates or simulated null distribution, constructed from pointwise CIs, as a graphical hypothesis test to reject clustering which amounts to multiple hypothesis testing and inflated type I errors (Figs. 2a & b) (Pollington et al., 2019a). Additionally they defined the *clustering endpoint D* as the distance at which the lower bound of the first pointwise (percentile) CI belonging to a point estimate above $\tau = 1$, touches $\tau = 1$ (Fig. 2a), or where the point estimate line touches the upper bound of the null envelope formed from pointwise CIs (Fig. 2b); thus mixing graphical hypothesis testing with point estimation (Pollington et al., 2019a).

*3.3. Graphical hypothesis test of no clustering and no inhibition*

We instead construct a *global envelope* around the distribution of the null hypothesis ($H_0$: $\tau = 1$, no spatiotemporal clustering nor inhibition) (Myllymäki and Mrkvička, 2019a). This is generated by randomly permuting the time marks (onset time$_i$) of the spatiotemporal data points $X_i =$ (x-coordinate$_i$, y-coordinate$_i$, onset time$_i$), $i = \{1, \ldots, n\}$ to scramble any spatiotemporal clustering present and simulate what $\hat{\tau}$ would be under $H_0$. We assess if a subset of distance bands $\underline{\delta}$ of $\underline{\Delta}$ exists (as contiguous or disjoint regions) where the tau point estimate $\hat{\tau}(d)$ is ever above/below the upper/lower bound, respectively, of this global null envelope. This global envelope is of extreme rank type ("defined as the minimum of pointwise ranks") with 95% significance level and extreme rank length *p*-value interval (note: a range, not a single *p*-value) as constructed by the GET *R* package (Myllymäki et al., 2019b) (see online Graphical abstract). We compute 2500 time-mark permuted tau simulations for an optimal test (Myllymäki et al., 2017).

The test is two-tailed, which is necessary as only once the graph is plotted is the presence of clustering or inhibition apparent (alternative hypothesis $H_1$: $\tau \neq 1$). One should use a *global* $\underline{\Delta}$ covering all pairwise distances, but at large distances null simulations commonly diverge from $\tau = 1$. This however can be assessed by tracking the upper & lower quartiles of null simulations (see code in Supplementary material) and choosing a shorter maximum distance; here the maximum pairwise distance is 319 m and we choose 220 m.

*3.4. Point estimation of the clustering endpoint and its sampling error*

If graphical hypothesis testing establishes evidence against no spatiotemporal clustering nor inhibition within a subset of distance bands $\underline{\delta}$ (Section 3.3) and visual graph inspection indicates clustering, then it is acceptable to estimate the *clustering endpoint* $\hat{D}$ for the clustering range [$d_1 = 0$ (assumed), $d_m = \hat{D}$); this is where the point estimate line $\hat{\tau}(d)$ intercepts $\tau = 1$, *i.e.* $\hat{\tau}(\hat{D}) = 1$. Due to discrete distance bands we linearly interpolate between their endpoints, *i.e.* [$d_l, d_m$) with $\hat{\tau} > 1$, and [$d_{l+1}, d_{m+1}$) with $\hat{\tau} < 1$, to obtain $\hat{D}$.
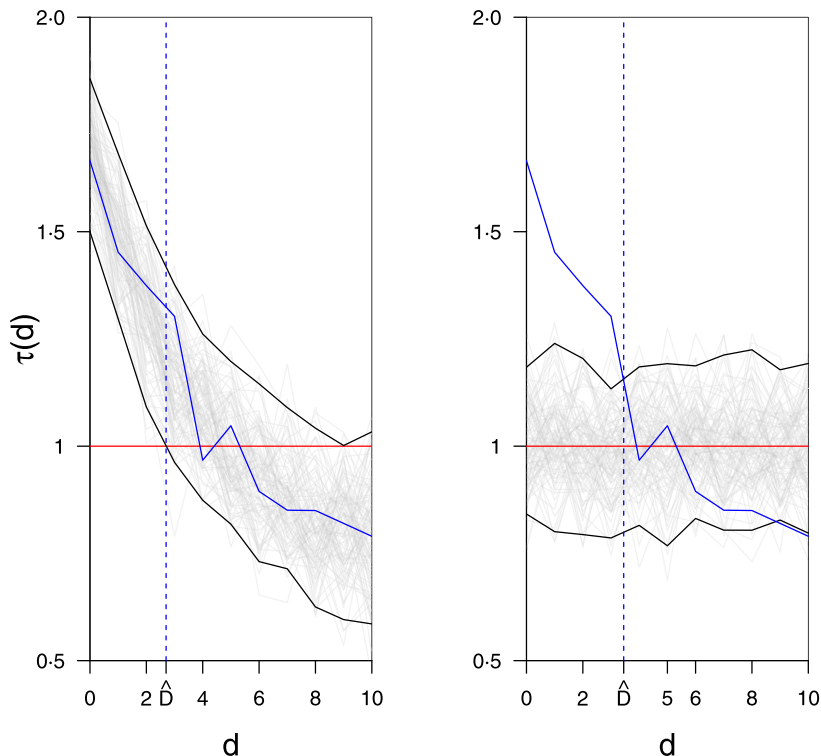
**Fig. 2.** Previous naive methods: several authors (Section 3.2) choose one envelope type as 'central' (a) or 'null' (b), then simultaneously test the hypothesis of no clustering and estimate the clustering endpoint parameter $\hat{D}$ (Pollington et al., 2019a). The single red line $\tau = 1$ represents no spatiotemporal clustering nor inhibition. Grey lines indicate (a) a collection of spatial bootstrap estimates $\hat{\tau}^*$ (denoted by $*$, see Section 3.4.1) from a typical tau estimator characterised by negative exponential lines with Normal noise, or (b) simulations of the tau estimator on time-mark permuted data for null envelope construction, represented here as lines at $\tau = 1$ with Normal noise; black lines mark out the envelope bounds constructed from pointwise confidence intervals. The solid blue line characterises an empirical tau point estimate $\hat{\tau}(d)$.

Instead, we split the method into the separate steps of graphical hypothesis testing (Fig. 4) and point estimation (*e.g.* Fig. 5) in Sections 3.3 & 3.4, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To estimate the sampling error of $\hat{D}$ we use spatial bootstrap (denoted by $*$, see Section 3.4.1) tau estimates $\underline{\hat{\tau}}^*$. For every bootstrap simulation (that represents a connected line of simulated tau estimates for increasing $d$, *i.e.* $\{\hat{\tau}^*(d_l, d_m) : [d_l, d_m) \in \underline{\Delta}\}$), we record those that originate from above $\tau = 1$ and then intersect $\tau = 1$ at some greater distance $D$ within $\underline{\Delta}$, *i.e.* those that satisfy $\{\hat{\tau}^*(D) = 1 : D \in \underline{\Delta}\}$. We use $N = 2500$ samples which is more than sufficient for a typical bootstrap (Efron and Tibshirani, 1998). We then take this horizontal set of values $\underline{D}$ and use it to obtain a CI to describe the sampling error in $\hat{D}$ (see online Graphical abstract).

We now investigate *spatial* bootstrap methods (Section 3.4.1), CI construction (Section 3.4.2) and distance band sets (Section 3.4.3).

### 3.4.1. Spatial bootstrap sampling methods for $\hat{\tau}$

The method that generates spatial bootstrap tau estimates $\hat{\tau}^*$ may impact the sampling error of the CI of the clustering endpoint $\hat{D}$. Through bootstrap theory, the sampling distribution $\underline{\hat{\tau}}^*$ may serve as a proxy for the actual distribution of $\hat{\tau}$ on the data. Furthermore the (central) envelopes
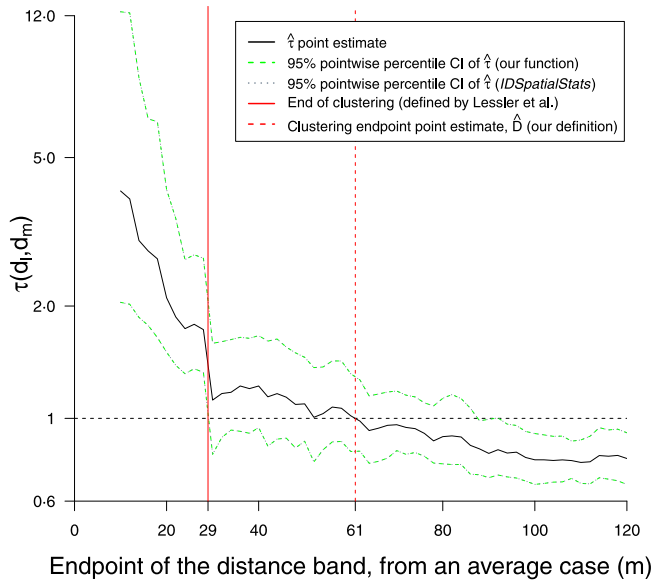
**Fig. 3.** Baseline result: a reproduction of a previous analysis (Lessler et al., 2016, Fig. 4C); note their x-axis used the midpoint of the distance band rather than the endpoint. The superimposition of their envelope and ours validates our implementation of tau functions from their IDSpatialStats *R* package (Giles et al., 2018). According to Lessler et al.'s convention, the end of the clustering range is where the lower bound of the envelope intersects $\tau = 1$ ($\hat{D}_{base} = 29$ m). We do *not* endorse this convention however, and take the point where the point estimate line $\hat{\tau}$ intersects ($\hat{D} = 61$ m). Additionally we generally recommend the use of BCa pointwise CIs for diagnostic plots such as this.
Their 29 m value was previously reported as 15 m due to a misinterpretation in the midpoint distance graph reading, as confirmed by Lessler (personal comm.)). Instead we use the endpoint so that a point $(d, \tau(d))$ read off the graph pertains to the single distance band $[d_{m-1}, d_m = d]$, and more commonly $[0, d)$ when $d_l = 0$. Whereas the midpoint requires an additional step by the reader that can easily introduce error.

constructed from $\hat{\underline{\tau}}^*$ may approximate the envelope of $\hat{\tau}$ on the data, and as a corollary the $\underline{D}$ envelope formed by the intercept of $\hat{\underline{\tau}}^*$ with $\tau = 1$ (Efron, 1979).

We compare three *spatial bootstrap* methods that define the bootstrap either on the data beforehand (see *Resampled-index spatial bootstrap* (RISB)), or bootstrap the local $\tau$ functions (see *Modified point spatial bootstrap* (MPSB) in Appendix A.4) or locally-(un/)related mark functions $m$ (see *Modified marked point spatial bootstrap* (MMPSB)). All are non-parametric because they randomly resample the data, or local $\tau$ or $m$ functions, without imposing a distribution (Loh, 2008).

*Resampled-index spatial bootstrap (RISB).* We start again with spatiotemporal data set $X = (X_i)_{i=\{1,...,n\}}$ where $X_i = $ (x-coordinate$_i$, y-coordinate$_i$, onset time$_i$). Using the Uniform distribution we resample with replacement the data's index vector $\underline{i} = (1, ..., n)$ $n$ times (equal to the number of cases), to generate a spatial bootstrap of the data, with sample index vector $\underline{i}^* = (i_1^*, ..., i_n^*)$ and subsequent data set $X^* = (X_{i*})$; $\underline{i}$ and $\underline{i}^*$ have the same length, but $\underline{i}^*$ is bound to contain duplicated indices due to sampling with replacement. To obtain $N$ bootstrap $\tau$ estimates $\hat{\underline{\tau}}^* = (\hat{\tau}_1^*, ..., \hat{\tau}_N^*)$ we apply the tau odds estimator individually to $N$ bootstrap data sets $X_1^*, ..., X_N^*$; the same approach could be applied to other $\tau$ estimators.

Loh critiques this "naive" sampling with replacement of the points $X_i$ of a spatial dataset $(X_i)$ to produce a spatial bootstrap sample, because "the spatial dependence structure has to be preserved as much as possible" (Loh, 2008) ... "to reflect properties of the original process" (Loh and Stein, 2004). Lessler et al. (2016) and others used this method and additionally for any $i^*$, $j^*$ resampled index pairs, dropped pairs $(X_{i*}, X_{j*})$ for which $X_{i*} = X_{j*}$, as they represented the same point, to avoid 'self comparisons'.

*Modified marked point spatial bootstrap (MMPSB).* Like MPSB (discussed in Appendix A.4) we bootstrap local functions not the data (RISB), but like all three methods still use $\underline{i}^*$ to decide the sample. Unlike MPSB rather than using local $\tau$ functions (Equation A.2), we go deeper and compute the number of locally-related/unrelated cases from mark functions $m_i(k)$ local to case $i$, according to their binary time-relatedness $k$.

The number of time-related cases (#related) out of all empirical cases $j$, within a distance $[d_l, d_m]$ around a case $i^*$ (chosen in the bootstrap sample) is:

$$\#\text{related}(d_l, d_m, k = 1, i^*) \equiv m_{i*}(d_l, d_m, k = 1) := \sum_{j \in j, j \neq i^*} \mathbb{1}(d_l \leq d_{i*j} < d_m, z_{i*j} = 1) \qquad (2)$$

and then an average is taken over the $n$ cases in the bootstrap sample of indices $\underline{i}^*$:

$$\overline{\#\text{related}^*(d_l, d_m)} \equiv m^*(k = 1) = \frac{1}{n} \sum_{i^* \in \underline{i}^*} \sum_{j \in j, j \neq i^*} \mathbb{1}(d_l \leq d_{i*j} < d_m, z_{i*j} = 1), \qquad (3)$$

and similar steps for time-unrelated cases yield:

$$\overline{\#\text{unrelated}^*(d_l, d_m)} \equiv m^*(k = 0) = \frac{1}{n} \sum_{i^* \in \underline{i}^*} \sum_{j \in j, j \neq i^*} \mathbb{1}(d_l \leq d_{i*j} < d_m, z_{i*j} = 0), \qquad (4)$$

and finally the odds and odds ratio estimators can be calculated as before:

$$\theta^*(d_l, d_m) = \frac{\overline{\#\text{related}^*(d_l, d_m)}}{\overline{\#\text{unrelated}^*(d_l, d_m)}} = \frac{\sum_{i^* \in \underline{i}^*} \sum_{j \in j, j \neq i^*} \mathbb{1}(d_l \leq d_{i*j} < d_m, z_{i*j} = 1)}{\sum_{i^* \in \underline{i}^*} \sum_{j \in j, j \neq i^*} \mathbb{1}(d_l \leq d_{i*j} < d_m, z_{i*j} = 0)} \qquad (5)$$

$$\tau^*_{\text{MMPSB}}(d_l, d_m) = \frac{\theta^*(d_l, d_m)}{\theta^*(0, \infty)} \qquad (6)$$

### 3.4.2. Confidence interval (CI) construction

Bias-corrected and accelerated (BCa) CIs can cope with asymmetrical distributions (like $\underline{D}$ defined in Section 3.4) better than percentile CIs. For non-parametric problems Carpenter and Bithell (2000) consistently found Efron's BCa method best, due to its low theoretical coverage errors for approximating the exact CI. BCa had "*second-order correct coverage*" errors under some assumptions, while a percentile CI was first-order correct at best (Efron, 1987). The BCa algorithm transforms a distribution of bootstrap calculations by normalisation to stabilise its variance so that a CI can be constructed, then back-transforms it (Efron, 1987). We calculated it using the coxed *R* package (Kropko and Harden, 2019).

### 3.4.3. Distance band sets

The tau statistic is non-unique as it depends on the distance band set chosen (Pollington et al., 2019a), so the potential variation in $\tau$ estimates from this choice is of interest; we explore this briefly in a non-systematic way. From analysing cases' pairwise distances we propose a reasonable distinct (non-overlapping) distance band set, i.e. $\underline{\Delta}_{\text{dis}} := \{[0,7), [7,15), [15,20), [20,25), [25,30), \ldots, [115,120 \text{ m})\}$ as a comparison to the overlapping set in the baseline analysis $\underline{\Delta}_{\text{overlap}} := \{[0,10), [0,12), [0,14), \ldots, [0,50), [2,52), [4,54), \ldots, [70,120 \text{ m})\}$ (Lessler et al., 2016), and test these using $N = 2500$ samples under MMPSB sampling.

## 4. Results & discussion

### 4.1. Dataset description

The epidemic over a small $\sim$280 x 240 m$^2$ area lasted nearly three months and five distinct generations can be discerned from the epidemic curve (Fig. B.1). Out of the 197 under-14 year olds, 185 became infected, along with three teenagers, leaving 377 remaining teenagers and adults
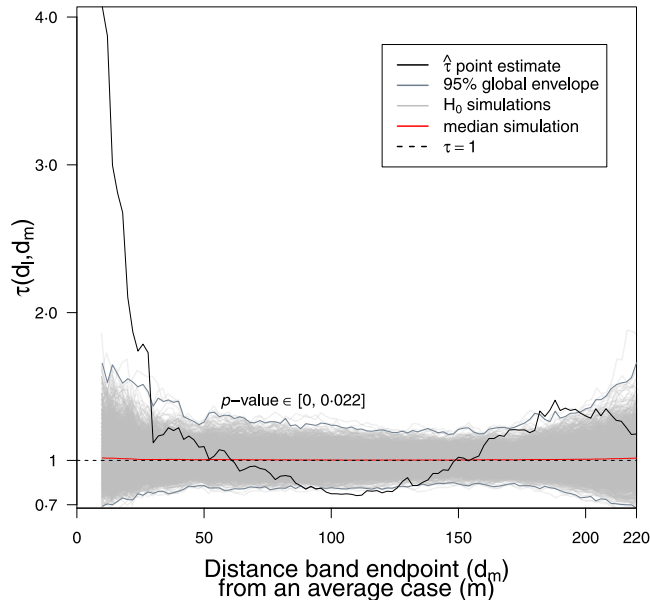
**Fig. 4.** Graphical hypothesis testing: Global envelope test, 'extreme rank' type, two-sided at 95% significance level using 2500 simulations of the null hypothesis ($H_0$: no spatiotemporal clustering nor inhibition, *i.e.* $\tau = 1$), p-value $\in [0, 0.022]$. Note there is a region where $\hat{\tau}$ just exits the global envelope lower bound (suggesting inhibition at $\sim$100 m) as well as the obvious departures above the upper bound (suggesting clustering at close distances and $\sim$190 m). We are confident that we are simulating $H_0$ well because the median simulation stays close to $\tau = 1$ throughout. Distance band set $:= \big\{[0, 10), [0, 12), [0, 14), \ldots, [0, 50), [2, 52), [4, 54), \ldots, [170, 220) \, \text{m}\big\}$.

uninfected (Neal and Roberts, 2004). Fig. B.2 indicates a weak signal of direct transmission between cases, as cases with onsets close together in time (shown by similar colours) tended to be nearby to each other.

### 4.2. Graphical hypothesis test: global envelope vs. pointwise CIs

There is moderate evidence against the hypothesis of no spatiotemporal clustering nor inhibition (p-value $\in [0, 0.022]$) based on constructing the global envelope around $\tau = 1$ under the null hypothesis (Fig. 4), and thus we conclude that the data $X$ is inconsistent with the null model ($H_0$: $\tau = 1$). So we turn to the alternative hypothesis, that there is clustering and/or inhibition. Fig. 4 suggests clustering at short distances and $\sim$190 m, and inhibition at long distances. Unfortunately it is not possible to compare our results with those of previous papers (see Section 3.2), since they used an incorrect pointwise CI approach to assess clustering, for which a p-value is not available.

### 4.3. Impact on the estimated clustering endpoint

Given the evidence of no clustering with the clear clustering signal in Fig. 4 we continue with parameter calculation. We estimate the clustering endpoint as $\hat{D} = 61.0$ m with a 95% percentile CI of $(29.0, 83.0 \, \text{m})$ over 100 bootstrap simulations using RISB sampling (Fig. 5), or $(29.2, 83.5 \, \text{m})$ over 2500 simulations (using 100% of simulations, see Appendix A.2); more bootstrap simulations do not appear to affect the sampling error.

The clustering identified at $\sim$190 m (Fig. 4) is ignored as we are interested in the first clustering range for baseline comparison; in general this may provide useful information of medium-range
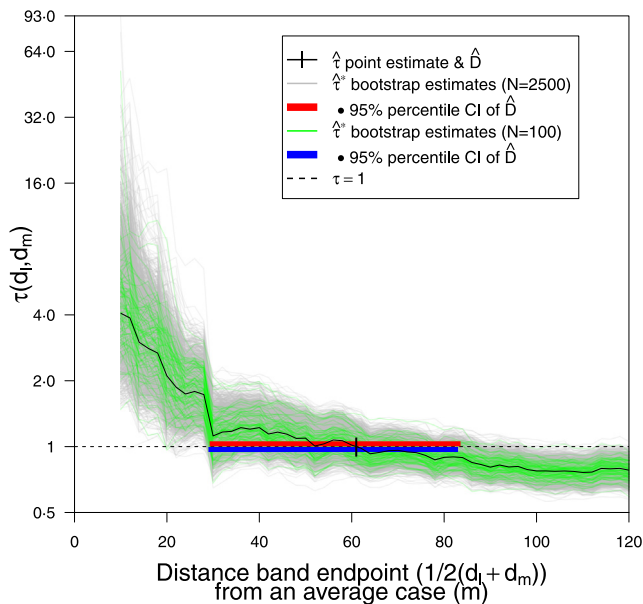
**Fig. 5.** Point estimation: Effect of number of samples on $\hat{D}$ sampling error, when using RISB sampling. Both CIs used 100% of simulations. $\hat{D} = 61\cdot0$ m; $N = 100$: 95% percentile CI (29·0, 83·0 m); $N = 2500$: CI (29·2, 83·5 m). Distance band set as Fig. 4.
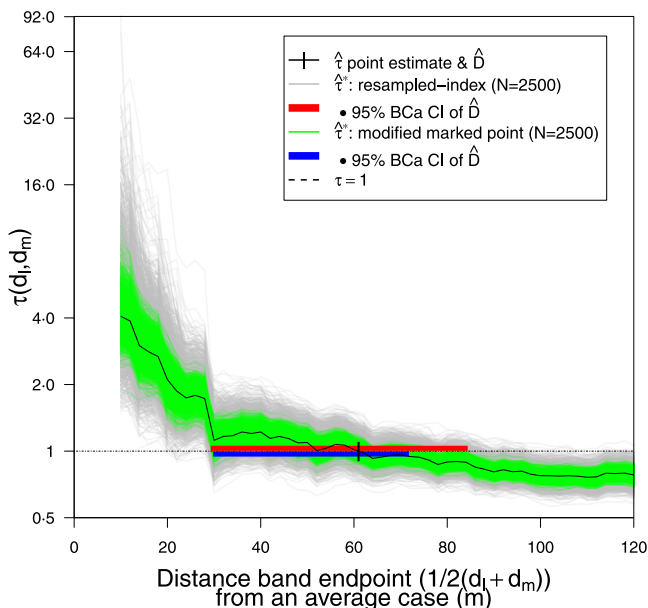


**Fig. 6.** Point estimation: Effect of spatial bootstrap sampling method on $\hat{D}$ sampling error. RISB 95% BCa CI (29·3, 84·4 m); MMPSB CI (29·8, 71·8 m); both CIs used 100% of simulations. Distance band set as Fig. 4, $\hat{D} = 61\cdot0$ m; $N = 2500$.
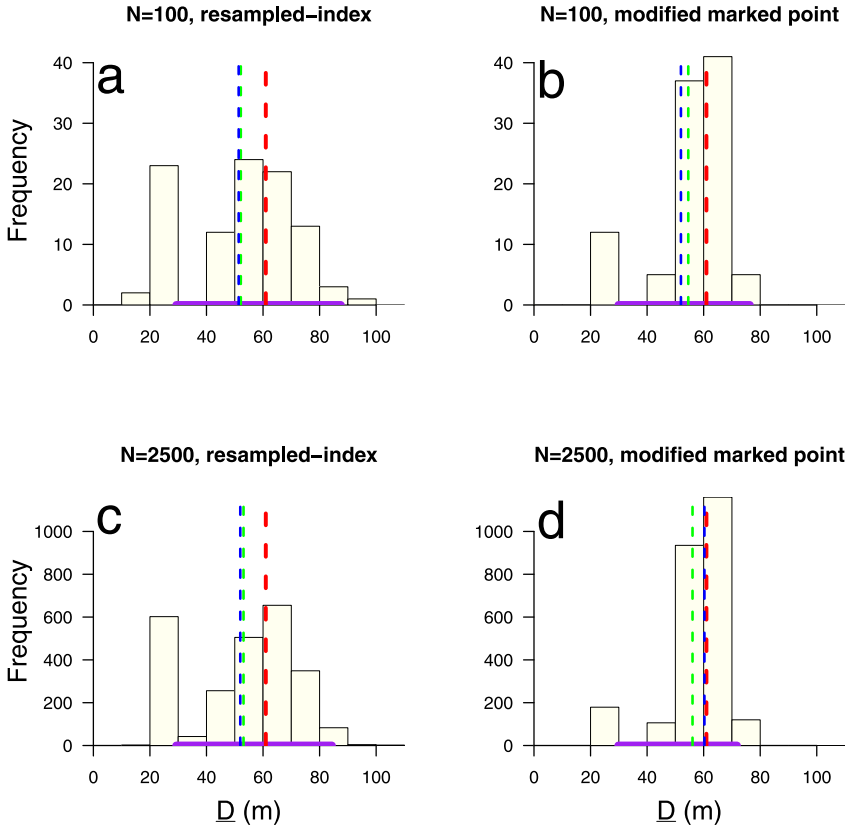
**Fig. 7.** Point estimation: The distribution of $\underline{D}$, formed by the sampled values of $\hat{D}_i$, i.e. $\underline{D} = \{\hat{D}_i : \hat{\tau}^*(\hat{D}_i) = 1, i = 1, \ldots, N\}$ (illustrated in online Graphical abstract), by number of bootstrap samples N = 100 (top row) or N = 2500 (bottom) and by spatial bootstrap sampling method RISB (left column) or MMPSB (right). Vertical dotted lines indicate the point estimate $\hat{D} = 61{\cdot}0$ m (red); with the mean (green) and median (blue) of the sampled estimates $\hat{D}_i$, obtained from where the bootstrap tau estimates $\hat{\underline{\tau}}^*$ intercept $\tau = 1$. RISB and MMPSB (N = 100) have positive skew as the mean estimate is greater than the median estimate, whereas for MMPSB (N = 2500) it has a negative skew. All bootstrap estimations have a negative bias with respect to mean or median summary measures versus the point estimate $\hat{D}$. At N = 2500 this is ~10 m for RISB and ~5 m for MMPSB. The data points used to construct the 95% BCa CIs (purple line on horizontal axis) from the $\hat{D}_i$ estimates in (a) are copied from Fig. 5 (N = 100 simulations) while those for (c) & (d) are from Fig. 6, while (b) has been freshly calculated. All four CIs used 100% of simulations. Distance band set as Fig. 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

spatial structure to understand disease spread, but secondary to control policy around an index case household.

The point estimate $\hat{D} = 61{\cdot}0$ m is 110% higher than the baseline clustering endpoint ($\hat{D}_{\text{base}} = 29$ m) (Fig. 3). Previous estimates derived via the improper method of finding the distance at which the lower bound of the central envelope (around $\hat{\tau}$) touches $\tau = 1$, have probably substantially underestimated this range. The plateauing shape of $\hat{\tau}(d)$ before it reaches $\tau = 1$ contributes to the increased imprecision in the estimate of $\hat{D}$. This highlights the utility of visually assessing the graph rather than rigidly using a $\tau = 1$ threshold, as it is likely that disease control over say a 60 m radius around an average case would see the biggest gains over the first 30 m with diminishing returns at wider radii (Fig. 6).
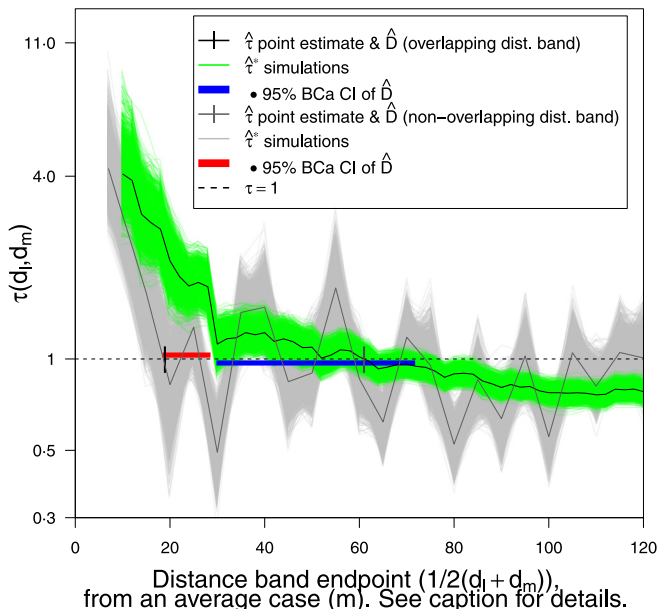
**Fig. 8.** Point estimation: Effect of distance band set on $\hat{D}$ sampling error using MMPSB sampling. Overlapping set $\underline{\Delta}_{\text{overlap}} := \{[0, 10), [0, 12), [0, 14), \ldots, [0, 50), [2, 52), [4, 54), \ldots, [70, 120)\, \text{m}\}$ (Lessler et al., 2016) and distinct $\underline{\Delta}_{\text{dis}} := \{[0, 7), [7, 15), [15, 20), [20, 25), [25, 30), \ldots, [115, 120)\, \text{m}\}$. $\underline{\Delta}_{\text{dis}}$ yields lower $\hat{D} = 18 \cdot 9$ m and more erratic point estimate $\hat{\tau}$ yet tighter 95% BCa CI for $\hat{D}$ ($18 \cdot 4$, $28 \cdot 6$ m) versus $\underline{\Delta}_{\text{overlap}}$ with $\hat{D} = 61 \cdot 0$ m and CI ($29 \cdot 8$, $71 \cdot 8$ m) however on further investigation the distribution $\underline{D}$ for the $\underline{\Delta}_{\text{dis}}$ is heavily bimodal; both CIs used 100% of simulations.

The 110% increase in the radial parameter $\hat{D}$ due to the corrected estimation method (Section 3.4) is further compounded for public health interventions, as their time and cost are more closely proportional to area not radius, and the areal increase is 342% (since $\pi(\hat{D}^2 - \hat{D}_{\text{base}}^2)/\pi\hat{D}_{\text{base}}^2 = 3 \cdot 42$, assuming $d_l = 0$).

## 4.4. Spatial bootstrap impact on sampling error: modified marked point vs. resampled-index

Using the modified marked point spatial bootstrap (MMPSB) (Section 3.4.1) yields a 24% narrower envelope than the resampled-index spatial bootstrap (RISB), leading to a 95% BCa CI for $\hat{D}$ of ($29 \cdot 8$, $71 \cdot 8$ m) (Fig. 6); both CIs used 100% of simulations.

If through simpler spatiotemporal structures or more evenly spread households the tau point estimate $\hat{\tau}$ had been smoother in approach to the $\tau = 1$ intercept, then the range of clustering would be far larger and the benefit of this estimation method (Section 3.4) more apparent. Given the reasons why MMPSB is better (Section 3.4.1), we believe RISB will poorly sample $\hat{D}$ and unfairly represent its precision.

MMPSB outperforms RISB because the latter loses more pair information from resampling indices and avoiding self-comparisons. This was checked empirically for the measles data: the tau point estimate is computed on 188 x 187 = 35156 pairs. In comparison, on average from 1000 simulations, RISB samples from 119 unique people, leading to 119 x 118 = 14042 unique pairs evaluated or $\sim 39 \cdot 9\%$ of the original pairs. Of course many additional duplicate pairs are used in RISB but we are only interested in unique pair information that is retained. MMPSB only has 119 unique mark functions, but each of them is compared with the other 187 cases, leading to 63·3% of pairs being retained.

*4.5. CI type: BCa vs. percentile*

Histograms of $\underline{D} = \{\hat{D}_i : \hat{\tau}^*(\hat{D}_i) = 1, i = 1, \ldots, N\}$ by number of bootstrap samples and sampling method strongly indicate asymmetric distributions, which support the principled use of BCa (Figs. 7a-d). Furthermore Figs. 7b & d show how MMPSB tau estimates $\hat{\underline{\tau}}^*$ intercept at values of $d$ closer to $\hat{D}$ than RISB. Despite this, percentile (29·2, 83·5 m) and BCa (29·3, 84·4 m) CIs, using RISB and $N = 2500$, differ marginally (Figs. 5 & 6). RISB appears to introduce a slight positive skew (mean > median) in $\underline{D}$ whereas MMPSB with sufficient samples ($N = 2500$) has a negative skew (Figs. 7a-d). At $N = 2500$, MMPSB noticeably reduces the bias ($\bar{D} - \hat{D}$), between mean/median estimates of $\underline{D}$ and the point estimate $\hat{D}$ from ~10 m to ~5 m, or ~8% of $\hat{D}$.

*4.6. Distance bands: overlapping vs. distinct*

Overlapping distance band sets $\underline{\Delta}_{\text{overlap}}$ appear to produce $\hat{D}$ estimates with higher variance (95% BCa CI (29·8, 71·8 m)) than distinct $\underline{\Delta}_{\text{dis}}$ (CI (18·4, 28·6 m)) (Fig. 8), but a clearer and smoother trend in tau with increasing distance (both CIs used 100% of simulations). The distribution of $\underline{D}$ is strongly bi-modal for $\underline{\Delta}_{\text{dis}}$ because the simulations are more erratic about $\tau = 1$. The increased volatility of $\hat{\tau}$ also results in multiple intercepts with $\tau = 1$, but for usability we prefer a single range of clustering, given in this case by $\underline{\Delta}_{\text{overlap}}$.

## 5. Conclusion and recommendations for improved use

We have shown that the way the endpoint of the clustering range $\hat{D}$ is currently calculated using the tau statistic can lead to biased estimates. Using a redefined $\underline{D}$ for the Hagelloch measles dataset, resulted in bias reductions equivalent to increasing the clustering area of elevated odds by 342%. An improved spatial bootstrap sampling method delivered $\hat{D}$ estimates with 24% lower sampling error. These improvements will appear in future versions of the IDSpatialStats R package (Giles et al., 2018). Our results (Section 4) support the following recommendations:

- using the point $d$ where the tau point estimate line $\hat{\tau}(d)$ intercepts $\tau = 1$ to define the clustering endpoint estimate $\hat{D}$ avoids underestimating it like earlier papers. However this estimation should be conditional on graphical hypothesis testing and visual plot inspection.
- the modified marked point spatial bootstrap should be used to simulate $\hat{\tau}$ instead of the resampled-index method, for CIs that better represent the precision of the clustering endpoint $\hat{D}$.
- BCa, rather than percentile, CIs should be used as they give better coverage since the distribution of bootstrap tau simulations $\hat{\underline{\tau}}^*$ or clustering endpoint estimates $\underline{D}$ is commonly asymmetric.

*Tau statistic limitations.* The distance band set choice $[d_l, d_m) \in \underline{\Delta}$ clearly biases $\hat{D}$ and affects its smoothness and sampling error. A better understanding of how to choose distance bands for a given purpose is now needed. It is also unknown how the time-relatedness interval choice $[T_1, T_2]$ (where $z_{ij} = \mathbb{1}((t_j - t_i) \in [T_1, T_2])$) biases the tau statistic through inclusion of extraneous co-primary or secondary cases. It is unclear how second-order correlation functions like the tau statistic and Ripley's $K$ function (Gabriel and Diggle, 2009), originally founded in spatiotemporal point processes with continuous support in $\mathbb{R}^2$, behave for this spatially discrete process. Finally the number of bootstrap samples required for graphical hypothesis testing and estimation purposes is unknown; we believe that related research by Davidson and MacKinnon (2000) could inform a heuristic algorithm.

<p align="center">* τ *</p>

We encourage the adoption of the statistical protocol described (see online Graphical abstract) to properly test for clustering, and, if appropriate, estimate its range. Control programmes have already been informed by the tau statistic and applying these bias-reduction methods will improve its accuracy in future health policy decisions. In addition to modellers or epidemiologists working on real-time outbreaks or post-study analysis, we hope statisticians are inspired to apply this statistic to spatiotemporal branching processes in new fields.

**CRediT contribution statement**

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgments**

**Figure summary**

The figures above illustrate the spatial coverage of the statistic (Fig. 1), show previous methods and the baseline analysis (Figs. 2a-b & 3), perform a graphical hypothesis test (Fig. 4), or investigate effects on the parameter point estimate $\hat{D}$ and distribution $\underline{D}$ (unless stated the distance band set is 'overlapping', see Fig. 4 caption):

- number of samples $N = 100$ or 2500, on RISB sampling using percentile CIs (Fig. 5)
- RISB *vs.* MMPSB, or MMPSB *vs.* MPSB sampling (Figs. 6 & B.3, respectively) (using $N = 2500$ and BCa CIs)
- RISB *vs.* MMPSB sampling and $N = 100$ or 2500 (with BCa CIs) (Figs. 7a-d)
- overlapping *vs.* distinct distance band sets (using $N = 2500$, MMPSB sampling and BCa CIs) (Fig. 8)

**Supplementary material**

Online supplementary material (https://doi.org/10.1016/j.spasta.2020.100438) contains Appendices A.1-5, detailing the computation methods and two of the spatial bootstrap sampling methods, while Appendix B contains Figs. B.1-3 that were not central to the manuscript. The analysis code in *R Markdown* is available from https://github.com/t-pollington/developments_tau_statistic under a GNU General Public License v3.0 licence.

# References

Azman, A.S., Luquero, F.J., Salje, H., Mbaïbardoum, N.N., Adalbert, N., Ali, M., Bertuzzo, E., Finger, F., Toure, B., Massing, L.A., Ramazani, R., Saga, B., Allan, M., Olson, D., Leglise, J., Porten, K., Lessler, J., 2018. Micro-hotspots of risk in urban cholera epidemics. J. Inf. Dis. 218 (7), 1164–1168. http://dx.doi.org/10.1093/infdis/jiy283.

Baddeley, A., Rubak, E., Turner, R., 2015. Spatial Point Patterns: Methodology and Applications with R, first ed. CRC Press/Taylor & Francis, Boca Raton, http://dx.doi.org/10.1201/b19708.

Bhoomiboonchoo, P., Gibbons, R.V., Huang, A., Yoon, I.-K., Buddhari, D., Nisalak, A., Chansatiporn, N., Thipayamongkolgul, M., Kalanarooj, S., Endy, T., Rothman, A.L., Srikiatkhachorn, A., Green, S., Mammen, M.P., Cummings, D.A.T., Salje, H., 2014. The spatial dynamics of dengue virus in Kamphaeng Phet, Thailand. PLoS Neglected Trop. Dis. 8 (9), 6–11. http://dx.doi.org/10.1371/journal.pntd.0003138.

Bland, J.M., 2000. An Introduction to Medical Statistics, third ed. Oxford medical publications, OUP Oxford, New York, USA.

Carpenter, J., Bithell, J., 2000. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat. Med. 19, 1141–1164. http://dx.doi.org/10.1002/(SICI)1097-0258(20000515)19:9<1141::AID-SIM479>3.0.CO;2-F.

Davidson, R., MacKinnon, J.G., 2000. Bootstrap tests: how many bootstraps?. Econometric Rev. 19 (1), 55–68. http://dx.doi.org/10.1080/07474930008800459.

Diggle, P.J., Kaimi, I, Abellana, R., 2010. Partial-likelihood analysis of spatio-temporal point-process data. Biometrics 66 (2), 347–354. http://dx.doi.org/10.1111/j.1541-0420.2009.01304.x.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. Ann. Statist. 7 (1), 1–26. http://dx.doi.org/10.1214/aos/1176344552.

Efron, B., 1987. Better bootstrap confidence intervals. J. Amer. Statist. Assoc. 82 (397), 171–185. http://dx.doi.org/10.1080/01621459.1987.10478410.

Efron, B., Tibshirani, R., 1998. An Introduction to the Bootstrap. Chapman & Hall/CRC, Boca Raton, London.

Gabriel, E., Diggle, P.J., 2009. Second-order analysis of inhomogeneous spatio-temporal point process data. Stat. Neerlandica 63 (1), 43–51. http://dx.doi.org/10.1111/j.1467-9574.2008.00407.x.

Giles, J., Salje, H., Lessler, J., 2018. IDSpatialStats *R* package development version v0.3.9. https://github.com/HopkinsIDD/IDSpatialStats.

Giles, J., Salje, H., Lessler, J., 2019. The IDSpatialStats R package: quantifying spatial dependence of infectious disease spread. The R Journal 11 (2), 308–327. https://journal.r-project.org/archive/2019/RJ-2019-043.

Grabowski, M.K., Lessler, J., Redd, A.D., Kagaayi, J., Laeyendecker, O., Ndyanabo, A., Nelson, M.I., Cummings, D.A.T., Baptiste Bwanika, J., Mueller, A.C., Reynolds, S.J., Munshaw, S., Ray, S.C., Lutalo, T., Manucci, J., Tobian, A.A.R., Chang, L.W., Beyrer, C., Jennings, J.M., Nalugoda, F., Serwadda, D., Wawer, M.J., Quinn, T.C., Gray, R.H., 2014. The role of viral introductions in sustaining community-based HIV epidemics in rural Uganda: Evidence from spatial clustering, phylogenetics, and egocentric transmission models. PLoS Med. 11 (3), http://dx.doi.org/10.1371/journal.pmed.1001610.

Grantz, K.H., Rane, M.S., Salje, H., Glass, G.E., Schachterle, S.E., Cummings, D.A.T., 2016. Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. Proc. Natl. Acad. Sci. USA 113 (48), 13839–13844. http://dx.doi.org/10.1073/pnas.1612838113.

Hoang Quoc, C., Salje, H., Rodriguez-Barraquer, I., In-Kyu, Y., Chau, N.V.V., Hung, N.T., Tuan, H.M., Lan, P.T., Willis, B., Nisalak, A., Kalayanarooj, S., Cummings, D.A.T., Simmons, C.P., 2016. Synchrony of dengue incidence in Ho Chi Minh City and Bangkok. PLoS Neglected Trop. Dis. 10 (12), 1–18. http://dx.doi.org/10.1371/journal.pntd.0005188.

Höhle, M., Meyer, S., Paul, M., Held, L., Burkom, H., Correa, T., Hofmann, M., Lang, C., Manitz, J., Riebler, A., Sabanés Bové, D., M., Salmon, Schumacher, D., Steiner, S., Virtanen, M., Wei, W., Wimmer, V., R Core Team, 2019. surveillance *R* package v1.18.0. https://cran.r-project.org/package=surveillance.

Kropko, J., Harden, J., 2019. coxed *R* package v0.3.2. https://cran.r-project.org/package=coxed.

Lessler, J., Salje, H., Grabowski, M.K., Cummings, D.A.T., 2016. Measuring spatial dependence for infectious disease epidemiology. PLoS One 11 (5), 1–13. http://dx.doi.org/10.1371/journal.pone.0155249.

Levy, J.W., Bhoomiboonchoo, P., Simasathien, S., Salje, H., Huang, A., Rangsin, R., Jarman, R.G., Fernandez, S., Klungthong, C., Hussem, K., Gibbons, R.V., Yoon, I.-K., 2015. Elevated transmission of upper respiratory illness among new recruits in military barracks in Thailand. Influenza Respir. Viruses 9 (6), 308–314. http://dx.doi.org/10.1111/irv.12345.

Loh, J.M., 2008. A valid and fast spatial bootstrap for correlation functions. Astrophys. J. 681 (1), 726–734. http://dx.doi.org/10.1086/588631.

Loh, J.M., Stein, M.L., 2004. Bootstrapping a spatial point process. Statist. Sinica 14 (1), 69–101.

Myllymäki, M., Mrkvička, T., 2019a. GET: Global envelopes in R. arXiv/stat.ME.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Hahn, U., Kuronen, M., Rost, M., Seijo, H., 2019b. GET *R* Package v0.1-6. https://cran.r-project.org/package=GET.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., Hahn, U., 2017. Global envelope tests for spatial processes. J. Royal Stat. Soc. Ser. B 79 (2), 381–404. http://dx.doi.org/10.1111/rssb.12172.

Neal, P.J., Roberts, G.O., 2004. Statistical inference and model selection for the 1861 Hagelloch measles epidemic. Biostatistics 5 (2), 249–261. http://dx.doi.org/10.1093/biostatistics/5.2.249.

Oesterle, H., 1992. Statistische Reanalyse Einer Masernepidemie 1861 in Hagelloch (PhD thesis). Eberhard-Karls-Universität Tübingen.

Pfeilsticker, A., 1863. Beiträge zur Pathologie Der Masern Mit Besonderer Berücksichtigung Der Statistischen Verhältnisse (PhD thesis). Eberhard-Karls-Universität Tübingen.

Pollington, T.M., Tildesley, M.J., Hollingsworth, T.D., Chapman, L.A.C., 2019a. The spatiotemporal tau statistic: a review. https://arxiv.org/abs/1911.11476.

Pollington, T.M., Tildesley, M.J., Hollingsworth, T.D., Chapman, L.A.C., 2019b. Epidemics[7] conference poster: Use global envelope tests not pointwise CIs, for graphical hypothesis tests of spatiotemporal clustering and tau statistic $\tau$. http://dx.doi.org/10.13140/RG.2.2.28985.52322.

R Core Team, 2019. *R* v3.6.3: A language and environment for statistical computing. https://www.r-project.org.

Rehman, N.A., Salje, H., Kraemer, M.U.G., Subramanian, L., Cauchemez, S., Saif, U., Chunara, R., 2018. Quantifying the impact of dengue containment activities using high-resolution observational data. http://dx.doi.org/10.1101/401653, bioRxiv.

Salje, H., Cauchemez, S., Alera, M.T., Rodriguez-Barraquer, I., Thaisomboonsuk, B., Srikiatkhachorn, A., Lago, C.B., Villa, D., Klungthong, C., Tac-An, I.A., Fernandez, S., Mark Velasco, J., Roque Jr, V.G., Nisalak, A., Macareo, L.R., Levy, J.W., Cummings, D.A.T., Yoon, I.-K., 2016a. Reconstruction of 60 years of chikungunya epidemiology in the Philippines demonstrates episodic and focal transmission. J. Inf. Dis. 213 (4), 604–610. http://dx.doi.org/10.1093/infdis/jiv470.

Salje, H., Cummings, D.A.T., Rodriguez-Barraquer, I., Katzelnick, L.C., Lessler, J., Klungthong, C., Thaisomboonsuk, B., Nisalak, A., Weg, A., Ellison, D., Macareo, L.R., Yoon, I.-K., Jarman, R., Thomas, S., Rothman, A.L., Endy, T., Cauchemez, S., 2018. Reconstruction of antibody dynamics and infection histories to evaluate dengue risk. Nature 557 (7707), 719–723. http://dx.doi.org/10.1038/s41586-018-0157-4.

Salje, H., Lessler, J., Berry, I.M., Melendrez, M.C., Endy, T., Kalayanarooj, S., A-Nuegoonpipat, A., Chanama, S., Sangkijporn, S., Klungthong, C., Thaisomboonsuk, B., Nisalak, A., Gibbons, R.V., Iamsirithaworn, S., Macareo, L.R., Yoon, I.-K., Sangarsang, A., Jarman, R.G., Cummings, D.A.T., 2017. Dengue diversity across spatial and temporal scales: Local structure and the effect of host population size. Science 355 (6331), 1302–1306. http://dx.doi.org/10.1126/science.aaj9384.

Salje, H., Lessler, J., Endy, T.P., Curriero, F.C., Gibbons, R.V., Nisalak, A., Nimmannitya, S., Kalayanarooj, S., Jarman, R.G., Thomas, S.J., Burke, D.S., Cummings, D.A.T., 2012. Revealing the microscale spatial signature of dengue transmission and immunity in an urban population. Proc. Natl. Acad. Sci. USA 109 (24), 9535–9538. http://dx.doi.org/10.1073/pnas.1120621109.

Salje, H., Lessler, J., Paul, K.K., Azman, A.S., Rahman, M.W., Rahman, M., Cummings, D.A.T., Gurley, E.S., Cauchemez, S., 2016b. How social structures, space, and behaviors shape the spread of infectious diseases using chikungunya as a case study. Proc. Natl. Acad. Sci. USA 113 (47), 13420–13425. http://dx.doi.org/10.1073/pnas.1611391113.

Succo, T., Noël, H., Nikolay, B., Maquart, M., Cochet, A., Leparc-Goffart, I., Catelinois, O., Salje, H., Pelat, C., de Crouy-Chanel, P., de Valk, H., Cauchemez, S., Rousseau, C., 2018. Dengue serosurvey after a 2-month long outbreak in Nîmes, France, 2015: was there more than met the eye?. Eurosurveillance 23 (23), http://dx.doi.org/10.2807/1560-7917.ES.2018.23.23.1700482.

Truelove, S.A., Graham, M., Moss, W.J., Metcalf, C.J.E., Ferrari, M.J., Lessler, J., 2019. Characterizing the impact of spatial clustering of susceptibility for measles elimination. Vaccine 37 (5), 732–741. http://dx.doi.org/10.1016/j.vaccine.2018.12.012.