

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Penttinen, Antti; Ylitalo, Anna-Kaisa

Title: Deducing self-interaction in eye movement data using sequential spatial point processes

Year: 2016

Version:

Please cite the original version:

Penttinen, A., & Ylitalo, A.-K. (2016). Deducing self-interaction in eye movement data using sequential spatial point processes. *Spatial Statistics*, 17, 1-21.
<https://doi.org/10.1016/j.spasta.2016.03.005>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Accepted Manuscript

Deducing self-interaction in eye movement data using sequential spatial point processes

Antti Penttinen, Anna-Kaisa Ylitalo

PII: S2211-6753(16)30002-1

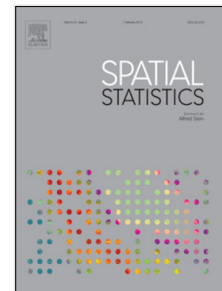
DOI: <http://dx.doi.org/10.1016/j.spasta.2016.03.005>

Reference: SPASTA 156

To appear in: *Spatial Statistics*

Received date: 30 October 2015

Accepted date: 28 March 2016



Please cite this article as: Penttinen, A., Ylitalo, A.-K., Deducing self-interaction in eye movement data using sequential spatial point processes. *Spatial Statistics* (2016), <http://dx.doi.org/10.1016/j.spasta.2016.03.005>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Deducing self-interaction in eye movement data using sequential spatial point processes

Antti Penttinen^a, Anna-Kaisa Ylitalo^{a,b,1,*}

^a*Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of
Jyväskylä, Finland*

^b*Department of Music, P.O. Box 35, FI-40014 University of Jyväskylä, Finland*

Abstract

24
25
26
27
28
29
30
31
32
33
34
35
36

Eye movement data are outputs of an analyser tracking the gaze when a person is inspecting a scene. These kind of data are of increasing importance in scientific research as well as in applications, e.g. in marketing and human-computer interface design. Thus the new areas of application call for advanced analysis tools. Our research objective is to suggest statistical modelling of eye movement sequences using sequential spatial point processes, which decomposes the variation in data into structural components having interpretation.

37
38
39
40
41
42
43
44
45
46

We consider three elements of an eye movement sequence: heterogeneity of the target space, contextuality between subsequent movements, and time-dependent behaviour describing self-interaction. We propose two model constructions. One is based on the history-dependent rejection of transitions in a random walk and the other makes use of a history-adapted kernel function penalized by user-defined geometric model characteristics. Both models are inhomogeneous self-interacting random walks. Statistical inference based on the likelihood is suggested, some experiments are carried out, and the models are used for determining the uncertainty of important data summaries for eye movement data.

47
48
49

Keywords: coverage, heterogeneous media, likelihood, recurrence, self-interacting random walk, stochastic geometry

50
51
52

*Corresponding author

Email address: anna-kaisa.ylitalo@jyu.fi (Anna-Kaisa Ylitalo)

¹Present address: University of Jyväskylä, Department of Music, P.O. Box 35, FI-40014 University of Jyväskylä, Finland

1. Introduction

Eye movements reflect brain functions, revealing information on ongoing cognitive processes, and can be recorded by eye trackers in a cost-efficient way. Eye movement data are spatio-temporal and consist of time sequences of *fixations*, points in the target space where the gaze stays for a while, and of *saccades*, which are rapid jumps between fixations. An example of eye movement data can be seen in Figure 1.

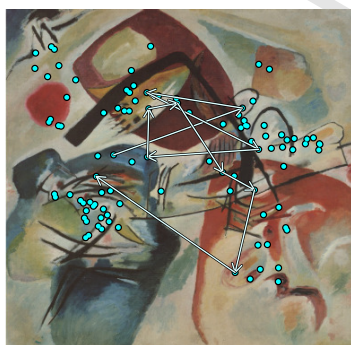


Figure 1: First 100 fixation points of one subject on a painting called Black Bow (1912) by Wassily Kandinsky. The arrows show the movement of the gaze during the first three seconds.

Fixation locations in eye movement data are point patterns which can be modelled by means of spatial point processes. Point process statistics is a well-developed branch of spatial statistics increasingly used in applied sciences, see e.g. Illian, Penttinen, Stoyan, and Stoyan (2008) and Diggle (2013). Extensive software `spatstat` (Baddeley, Rubak, and Turner, 2015) has made efficient point pattern data analysis attractive. Point process statistics has been applied for eye movement data by Barthelmé, Trukenbrod, Engbert, and Wichmann (2013), who use the spatial inhomogeneous Poisson point process to predict the fixation locations. The approach by Barthelmé et al. (2013) aggregates the eye movement data over time but omitting all dynamics. Engbert, Trukenbrod, Barthelmé, and Wichmann (2015) present a dynamical model that takes spatial interaction into account, but their model validation is based on characteristics of spatial point processes. A step towards a dynamic model is to add the temporal order of fixations, which leads us to the class of (finite) sequential spatial point processes, see van Lieshout (2006a,b, 2009). If, in addition to the order, the time instances of occurrences of the points are recorded and included in the model, then the underlying process is a spatio-temporal point process, see e.g. Diggle, Kaimi, and Abellana (2010), and an application to eye movement data by Ylitalo, Särkkä, and Guttorp (in press).

1
2
3
4
5
6
7
8
9 We consider eye movement data to be a realisation of a sequential spatial point
10 process which allows us to extend the approach by Barthelmé et al. (2013) for
11 detecting new important dynamic structures of data. The advantage of this
12 approach is that the likelihood is tractable and the simulation of realisations is
13 straightforward. In addition, sequential point process modelling is a construc-
14 tion step for spatio-temporal point processes.

15
16 For eye movement data, three structural components of sequential spatial point
17 processes are central. Spatial *heterogeneity* of fixation pattern means that some
18 parts of the scene get the observer's attention more than others. This strong
19 component is present in almost all eye movement data. It is usually modelled
20 through a saliency map, which is calculated from the features of the scene (see
21 e.g. Itti and Koch, 2000; Kümmerer, Wallis, and Bethge, 2014), such that the
22 most salient areas are expected to obtain more fixations. *Dynamic contextual-*
23 *ity* is a saccadic property which describes the (metric) length of a jump from
24 the current fixation to the next one: for example, nearby sites may be more
25 favourable than the more distant ones (see e.g. Tatler, Baddeley, and Vincent,
26 2006). Both heterogeneity and contextuality are well-established in eye move-
27 ment studies (see e.g. Barthelmé et al., 2013; Engbert et al., 2015; Kümmerer
28 et al., 2014).

29
30
31 However, our empirical evidence shows that these two components are not suffi-
32 cient, since e.g. they cannot model the learning effect. Similar findings are made
33 by Engbert et al. (2015) and Kümmerer et al. (2014). Thus we are looking for
34 simple mechanisms which could utilize the long-term dependence indicating the
35 learning process during an experiment. One such possibility is *self-interaction*,
36 which modifies the individual moves (saccades) by means of the history of the
37 eye movement sequence. As an illustration, the observer prefers to inspect the
38 whole scene at the beginning of the experiment and gradually focuses on a few
39 details (see e.g. Locher, Gray, and Nodine, 1996). Here, we offer a tool for
40 studying the self-interaction effect in eye movement sequences. We suggest new
41 models for eye movement data to deduce the effect of structural components and
42 to evaluate statistical variation in problem-specific functional summary statis-
43 tics. The suggested models have potential use also beyond the eye movement
44 research, e.g. in ecology for modelling animal movements, and in user-interface
45 studies.

46
47
48 Our idea is that the history of the sequence changes the dynamics during the
49 evolution of the eye movement sequence. We present two general principles for
50 model construction, both of which are generalisations of the random walk in
51 heterogeneous media. The first principle is the history-dependent thinning of
52 transitions, which assigns smaller weights for suggested transitions being at odds
53 with the chosen functional summary characteristic conditional on all previous
54 fixations. This type of penalization is similar to the area-interaction process
55 by Baddeley and van Lieshout (1995) in point process statistics. The second
56 principle resembles the ARCH (autoregressive conditional heterogeneity) model,
57
58

commonly applied in econometric time series analysis (Engle, 1982). Based on these principles, we present the construction of the two processes, how to simulate them, and how to estimate the parameters by the maximum likelihood method. Several summary statistics, assisted by Monte Carlo simulation, are applied in model evaluation.

The new models are mainly of “statistical” nature, which means that they do not mimic the neural process, but they can capture essential variation in eye movement data. We will not employ all the generality the suggested new models are able to achieve. Instead, the objective is to present the new ideas in terms of rather simple models which still are useful in eye movement data analysis, especially in the study of the learning mechanism during an experiment, and in the derivation of statistical variation of important data summaries. Furthermore, this new approach will bring eye movement data analysis closer to statistical inference. The motivation is the complexity of eye movement data, which are inhomogeneous in space and time, and the use of asymptotic inference, for example, is difficult to justify.

The paper is organized in the following way. In Section 2, two new sequential spatial point process models are suggested. Simulation algorithms are given in Section 3 with simulation experiments demonstrating the models. In Section 4, eye movement data in the field of art study is modelled by using the new approach to deduce self-interaction. Section 5 contains some concluding remarks.

2. Finite sequential spatial point process models

Suppose $\vec{x}_n = (x_1, \dots, x_n)$ is a sequence of time-ordered points in a bounded window $W \subset \mathbb{R}^2$. The corresponding unordered point set $\{x_1, \dots, x_n\}$ is denoted by $\{\vec{x}_n\}$. If (W^n, \mathcal{W}^n) stands for the n -dimensional space of ordered points provided with the Borel σ -algebra in W^n , the density function f w.r.t. the Lebesgue measure is defined sequentially as follows: $f_1(x_1)$ stands for the probability density of the first point, and the conditional density of a further point x given $\vec{x}_k = (x_1, \dots, x_k)$, $k = 1, \dots, (n-1)$, is denoted by $f_{k+1}(x|\vec{x}_k)$, a transition probability density for the transition $\vec{x}_k \rightarrow (\vec{x}_k, x)$, and the joint density of \vec{x}_n is

$$f(\vec{x}_n) = f_1(x_1) \prod_{k=1}^{n-1} f_{k+1}(x_{k+1}|\vec{x}_k). \quad (1)$$

The density $f_1(x)$ can be assumed to follow a function (e.g. the saliency map) exposing the focal areas of the target, or modelling can be conditional on the first observation x_1 . The transition densities $f_{k+1}(x|\vec{x}_k)$, $k = 1, \dots, n-1$, reflect saccadic features of the eye movement sequence.

A simple model for the transition would be a random walk in heterogeneous media, which is defined as

$$f_{k+1}(x|\vec{x}_k) \propto \alpha(x) K(x_k, x), \quad (2)$$

where $\alpha(x)$ is non-negative and bounded in W , and $K(x_k, x)$ is a Markovian kernel, i. e.,

$$\begin{aligned} K(x_k, x) &\geq 0 \text{ for all } x_k, x \in W, \text{ and} \\ \int_W K(x_k, u) du &= 1 \text{ for all } x_k \in W, \end{aligned}$$

$k = 1, \dots, n - 1$. In this simple model, $\alpha(x)$ describes heterogeneity of the scene. It can be a known saliency map, an empirical saliency map estimated as the intensity of repeated fixation patterns (see the discussion in Diggle, Gómez-Rubio, Brown, Chetwynd, and Gooding, 2007), or a model based prediction of the saliency map, for example,

$$\alpha(x) = h \left(\sum_{j=1}^p b_j z_j(x) \right),$$

where the variables $z_j(x)$ are the values of p feature vectors at x extracted from the scene using machine learning techniques, b_j 's are regression coefficients, and h is an adequate non-negative function (see e.g. Barthelmé et al., 2013). The Markovian kernel $K(x_k, x)$ describes the contextuality of subsequent fixations related to jump lengths. An example is the truncated Gaussian kernel

$$K(x_k, x) \propto e^{-\frac{1}{2\sigma^2} \|x_k - x\|^2}, \quad x_k, x \in W, \quad (3)$$

where $\|x_k - x\|$ is the Euclidean distance between the points x_k and x . For the rectangular window $W = [a, b] \times [c, d]$ the normalization term for (3) can be written as

$$2\pi\sigma^2(\Phi(b) - \Phi(a))(\Phi(d) - \Phi(c)),$$

where Φ is the c.d.f. of the standard normal distribution. This kernel penalizes large jumps and keeps the process inside the specified window W . Thus the model (2) captures heterogeneity in the target and models the transitions (or saccades) in a Markovian way.

The transition mechanism (2) may be insufficient if data contain learning. For instance, a two-stage model describing the nature of an aesthetic experience (see e.g. Locher, 2006; Locher et al., 1996, 2007) suggests that a picture is first inspected globally and, after having obtained a gist of the scene, the viewer starts to concentrate on some details. Also, the visual information gathered from the scene affects our cognitive processes and attention, which again affect the movement of the gaze. By keeping these complexities of human attention in mind, we develop two models which try to catch the sequential adaptation in the eye movement sequence in a tractable manner using geometric reasoning.

2.1. Self-interaction due to history-dependent rejection model

First, we define a *history-dependent rejection model* (later: rejection model), in which the self-interaction mechanism is created by a reweighting probability function. This model penalizes the location of the next point in terms of *coverage* or *recurrence* composed by the previous points: the density for the transition $\vec{x}_k \rightarrow (\vec{x}_k, x)$ is assumed to be

$$f_{k+1}(x|\vec{x}_k) \propto \alpha(x) K(x_k, x) \pi(x, S(\vec{x}_k, x)), \quad (4)$$

where $\pi(x, S(\vec{x}_k, x))$ is the reweighting probability of x when proposed according to the density proportional to $\alpha(x)K(x_k, x)$. Here, $S(\vec{x}_k, x) = S(x_1, \dots, x_k, x)$ is a measure of coverage or recurrence of the ordered sequence (x_1, \dots, x_k, x) .

Reasonable choices of the reweighting probability π in the eye movement context are given below.

Coverage-based reweighting

Two measures of the coverage of a point set are the area of its convex hull and the area of the associated ball union. From now on we assume that the scene W is convex. The convex hull of a point set $\{\vec{x}_k\}$, denoted by $\text{Conv}(\vec{x}_k)$, is the minimal convex subset of W which contains all the points of $\{\vec{x}_k\}$. The convex hull is unique and invariant under permutation of the points; hence it is the same for ordered and unordered sets. Let us denote by $S_C(\vec{x}_k)$ the area of $\text{Conv}(\vec{x}_k)$ and call it convex hull coverage.

The ball union measure of a point set $\{\vec{x}_k\}$ is defined as

$$\text{Bcov}(\vec{x}_k) = \bigcup_{i=1}^k b(x_i, r) \cap W,$$

where $b(x, r)$ stands for the ball with radius r and centred at x . It is a “regionalized” version of the point set, where the r close neighbourhood of a point is taken into account, and which again is invariant under permutation. Its area $S_B(\vec{x}_k)$ is called ball union coverage.

The rationale behind model (4) is that the kernel function generates random jumps and the reweighting probability determines which of the proposed jumps are accepted, depending on the current coverage of the sequence and on the new suggestion. Consider the convex hull coverage first: if the new suggestion x is not in $\text{Conv}(\vec{x}_k)$, the odds ratio of acceptance w.r.t. a proposal $y \in \text{Conv}(\vec{x}_k)$ with $\|x - x_k\| = \|y - x_k\|$ is $S_C(\vec{x}_k, x)/S_C(\vec{x}_k)$.

A reasonable and simple choice of the geometric nature for the reweighting probability would be

$$\pi(x, S(\vec{x}_k, x)) = \begin{cases} 1 & \text{if } x \in W \setminus \text{Conv}(\vec{x}_k) \\ \rho & \text{if } x \in \text{Conv}(\vec{x}_k) \end{cases}, \quad (5)$$

where $\rho \in [0, 1]$ and $k \geq 1$. If $\rho = 1$, we have the random walk model. When $\rho < 1$, this choice encourages locations outside the convex hull of previous points leading to faster coverage. When the convex hull of the points covers almost the whole scene, the process behaves like a random walk. The density for the transition $\vec{x}_k \rightarrow (\vec{x}_k, x)$ with the truncated Gaussian kernel is

$$f_{k+1}(x|\vec{x}_k) \propto \alpha(x) e^{-\frac{1}{2\sigma^2} \|x_k - x\|^2} (\mathbf{1}_{W \setminus \text{Conv}(\vec{x}_k)}(x) + \rho \mathbf{1}_{\text{Conv}(\vec{x}_k)}(x)), \quad (6)$$

where $k = 1, 2, \dots, n-1$, and $\mathbf{1}(\cdot)$ is the indicator function.

The convex hull coverage can be replaced by the ball union coverage. It is not as sensitive to distant points as the convex hull coverage but reacts to the ‘‘holes’’ in the point pattern. If $\text{Conv}(\vec{x}_k)$ is replaced by $\text{Bcov}(\vec{x}_k)$ in the reweighting probability (6), the process favours locations away from the previous points and hence reduces clustering if ρ is small. It should be noted that the ball union coverage measure requires the radius of the ball.

Recurrence-based reweighting

As a measure of recurrence we propose the number of earlier visits in a ball $b(x, r)$ around a site x , formally $\tilde{S}_R(\vec{x}_k, x) = \sum_{i=1}^k \mathbf{1}_{b(x_i, r)}(x)$. However, instead of using all the previous points, at step k the number of earlier visits is calculated from the point set $\{\vec{x}_{k-1}\}$ omitting the most recent point x_k . This delayed recurrence measure

$$S_R(\vec{x}_k, x) = \sum_{i=1}^{k-1} \mathbf{1}_{b(x_i, r)}(x) \quad (7)$$

is less confounded with the Markovian kernel $K(x_k, x)$ than $\tilde{S}_R(\vec{x}_k, x)$ and is therefore used in this paper from now on. Note also that the recurrence measure is not invariant under random permutation.

A simple model for the reweighting probability is

$$\pi(x, S_R(\vec{x}_k, x)) = \begin{cases} \theta & \text{if } S_R(\vec{x}_k, x) \geq 1 \\ 1 - \theta & \text{if } S_R(\vec{x}_k, x) = 0 \end{cases}, \quad (8)$$

where $\theta \in [0, 1]$ and $k \geq 2$. The odds ratio for accepting a location close to the points of $\{\vec{x}_{k-1}\}$ against accepting a location from an empty area is $\theta/(1 - \theta)$. If θ is close to 1, the process favours clustering, and if θ is small, the process avoids previously visited local areas around the points $\{\vec{x}_{k-1}\}$. If $\theta = 0.5$, the process is a random walk. The density for the transition $\vec{x}_k \rightarrow (\vec{x}_k, x)$ with the truncated Gaussian kernel is

$$f_{k+1}(x|\vec{x}_k) \propto \alpha(x) e^{-\frac{1}{2\sigma^2} \|x_k - x\|^2} ((1 - \theta) \mathbf{1}_{\{S_R(\vec{x}_k, x)=0\}}(x) + \theta \mathbf{1}_{\{S_R(\vec{x}_k, x) \geq 1\}}(x)), \quad (9)$$

where $k = 2, 3, \dots, n - 1$, (and $f_2(x|\vec{x}_1) = f_1(x)$).

In particular, the model defined through (4) is among the simplest ones which satisfy our requirements of self-interacting nature. Note that we need the normalized transition kernel in the likelihood, because the scaling factor contains the parameters of the model. The normalizing integral can be computed using numerical integration. Its evaluation can be avoided in the simulation of the process, however.

2.2. Self-interaction due to history-adapted model

The motivation behind the *history-adapted model* arises from the saliency map idea and the two-stage model by Locher and colleagues (Locher, 2006; Locher et al., 1996, 2007). Heterogeneity of the target plays the main role at an early stage of the process evolution: the areas with high saliency (or intensity) get more fixations than the low-saliency areas. However, when the target has been inspected well enough, the jump lengths get shorter as if the process were mimicking a local inspection process.

This model construction is intended for coverage type self-interaction. We apply directly an adaptive Markovian kernel $K_{\phi_k}(x_k, x)$, where ϕ_k is a function of the points $\{\vec{x}_k\}$ and determines the width of the kernel. Hence, the kernel changes in time and affects the jump lengths. The transition probability density can be written as

$$f_{k+1}(x|\vec{x}_k) = \frac{\alpha(x) K_{\phi_k}(x_k, x)}{\int_W \alpha(u) K_{\phi_k}(x_k, u) du} \quad (10)$$

$$\phi_k = \phi_k(\vec{x}_k) \propto H(S(\vec{x}_k)) \quad (11)$$

where $H(s)$ is decreasing in s . Here $\alpha(x)$ controls the heterogeneity of the target as in the rejection model, whilst $H(s)$ models the progress of the coverage. This model resembles the autoregressive conditional heterogeneity model (ARCH) commonly applied in time series analysis for modelling volatility (Engle, 1982). While ARCH models use the information from q lagged values, our model is allowed to use the entire history. Again, this model is a self-interacting random walk.

In what follows, we make use of the specific model

$$K_{\phi_k}(x_k, x) \propto e^{-\frac{1}{2\phi_k(\vec{x}_k)} \|x_k - x\|^2}, \quad (12)$$

$$\phi_k(\vec{x}_k) = \tau e^{-\kappa S(\vec{x}_k)/|W|}, \quad (13)$$

$\tau, \kappa \geq 0$ and $x_k, x \in W$, where $S(\vec{x}_k)$ is the coverage of $\{\vec{x}_k\}$. If $\kappa = 0$, the process is a random walk since the kernel does not change in time. This model contains two parameters, τ describing the initial kernel width and κ modelling the decay as a function of coverage. The transition is determined by the conditional density (10). Both convex hull and ball union coverages are suitable for this construction.

2.3. Model fitting and statistical inference

2.3.1. Model fitting

We assume that an ordered sequence $\vec{x}_n = (x_1, \dots, x_n)$ is observed in W . In what follows we suggest parameter estimation for the two models defined by (6) (rejection model), and by (12) and (13) (history-adapted model) assuming that the non-negative heterogeneity component $\alpha(x)$ is fixed. In practice, the estimation of $\alpha(x)$ is problematic, as we have pointed out in Discussion.

History-dependent rejection model

The log-likelihood for the general rejection model (4) is now a function of the model parameters. For the two parameter model defined by (6) the expression

$$\begin{aligned}
 l(\sigma^2, \rho) &= \sum_{k=1}^{n-1} \log(\alpha(x_{k+1})) - \frac{1}{2\sigma^2} \sum_{k=1}^{n-1} \|x_k - x_{k+1}\|^2 \\
 &+ \log(\rho) \sum_{k=1}^{n-1} 1_{\text{Conv}(\vec{x}_k)}(x_{k+1}) \\
 &- \sum_{k=1}^{n-1} \log \int_W \alpha(u) e^{-\frac{1}{2\sigma^2} \|x_k - u\|^2} (1_{W \setminus \text{Conv}(\vec{x}_k)}(u) + \rho 1_{\text{Conv}(\vec{x}_k)}(u)) du
 \end{aligned} \tag{14}$$

is obtained. Here we use the convex hull coverage in the reweighting probability, but also the ball union coverage could be used. The log-likelihood function for the rejection model with recurrence (9) is shown in Section 4.1, formula (16). The logarithm of the normalizing factor (the last line of (14)) can be computed by numerical integration. The optimization of $l(\sigma^2, \rho)$ w.r.t. σ^2 can be conducted using numerical optimization, or alternatively, one can solve the exponential family likelihood equation

$$\sum_{k=1}^{n-1} \mathbf{E}_{\sigma^2, \rho} (\|x_k - U\|^2 | \vec{x}_k) = \sum_{k=1}^{n-1} \|x_k - x_{k+1}\|^2,$$

where the expectation is over the conditional distribution of a new random point U from the distribution $f_{k+1}(x | \vec{x}_k)$. This can be computed using Monte Carlo maximum likelihood (MCMCML, see Geyer (1991)).

Maximizing the log-likelihood (14) (or (16)) is costly due to the normalizing integral. We apply the coordinate ascent algorithm which would be simple choice also for more complex model.

History-adapted model

The kernel width ϕ_k of the k th transition of the random walk is a function of the model parameters and is adapted to the history of the sequence through the coverage measure $S(\vec{x}_k)$. The log-likelihood for the general model given by (10) and (11) is

$$\sum_{k=1}^{n-1} \left[\log(\alpha(x_{k+1})) + \log K_{\phi_k}(x_k, x_{k+1}) - \log \int_W \alpha(u) K_{\phi_k}(x_k, u) du \right].$$

In the special case of (12) and (13) the log-likelihood is

$$\begin{aligned} l(\tau, \kappa) &= \sum_{k=2}^{n-1} \log(\alpha(x_{k+1})) - \sum_{k=2}^{n-1} \frac{1}{2\phi(\tau, \kappa)} \|x_k - x_{k+1}\|^2 \\ &\quad - \sum_{k=2}^{n-1} \log \int_W \alpha(u) e^{-\frac{1}{2\phi(\tau, \kappa)} \|x_k - u\|^2} du \end{aligned} \quad (15)$$

with $\phi_k(\kappa, \tau) = \tau e^{-\kappa S(\vec{x}_k)/|W|}$, $k = 1, \dots, n-1$.

The log-likelihood can again be maximized directly, or alternatively, the likelihood equations can be derived and solved: the estimation equations are

$$\begin{aligned} \sum_{k=2}^{n-1} \mathbf{E}_{\tau, \kappa} (\|x_k - U\|^2 | \vec{x}_k) / \phi_k &= \sum_{k=2}^{n-1} \|x_k - x_{k+1}\|^2 / \phi_k \\ \sum_{k=2}^{n-1} \mathbf{E}_{\tau, \kappa} (S(\vec{x}_k) \|x_k - U\|^2 | \vec{x}_k) / \phi_k &= \sum_{k=2}^{n-1} S(\vec{x}_k) \|x_k - x_{k+1}\|^2 / \phi_k, \end{aligned}$$

where the expectations are over the conditional distribution $f_{k+1}(x | \vec{x}_k)$ with parameters τ and κ . The estimation equations are in accordance with the maximum likelihood equations for the exponential family of distributions.

2.3.2. Model evaluation

Model evaluation of spatial dynamic models is typically based on the selected functional summary statistics which measure different features of the model, such as coverage, recurrence and jump length as a function of time (or order). In addition, a saliency map is plotted together with the fixation locations. The random variation of the summary statistics is estimated from simulations.

Model evaluation is done by estimating several summary statistics from data and plotting the estimates together with the model based simulated pointwise envelopes being a parametric bootstrap method (see e.g. Efron and Tibshirani, 1994, p. 53). These envelopes indicate statistical variation in the summary statistic under the parametric model assumption. It should, however, be noted

that when using the pointwise envelopes as statistical tests, the multiple testing problem is present and the interpretation of the envelopes must be done with care, see the discussions in Grabarnik, Myllymäki, and Stoyan (2011) and in Baddeley, Diggle, Hardegen, Lawrence, Milne, and Nair (2014).

Model evaluation is illustrated in the examples in Sections 3 and 4.

3. Simulation experiments

Realisations from the suggested models can be simulated sequentially using conditional distributions (4) and (10)-(11). We recommend to use the scaled heterogeneity $\alpha(x)/\max_{u \in W} \alpha(u)$ as the distribution for the first location, or alternatively, to condition to the first location x_1 of data. Assume that (x_1, \dots, x_k) are simulated. A simple algorithm for adding a point x to $\vec{x}_k = (x_1, \dots, x_k)$, or equivalently, simulating from the distribution having density $f_{k+1}(x|\vec{x}_k)$, is to apply the accept-reject algorithm, see e.g. Ripley (1987, p. 61), which provides an upper bound for the conditional density. For the two models, the algorithm is as follows:

- History-dependent rejection model: A point x following the conditional density $\alpha(x)K(x_k, x)$ is proposed using the accept-reject method and the proposal is accepted with the reweighting probability $\pi(x, S(\vec{x}_k, x))$.
- History-adapted model: The kernel width $\phi_k = H(S(\vec{x}_k))$ is computed and proposals from the unnormalized transition density $\alpha(x)K_{\phi_k}(x_k, x)$ are drawn using the accept-reject method.

In the following simulation experiment we generate realisations of the new models with three parameter values in order to demonstrate the time evolution of the new processes and to see how data summaries capture their properties. We illustrate to what extent and how fast these processes can cover the target area, as well as whether the process starts to cluster in space. Each realisation consists of 100 points located in the unit square window. In this illustration, for the sake of simplicity, we assume that the target space is homogeneous, setting $\alpha(x) \equiv 1$, and hence the first point is drawn uniformly and the same starting point is used for all realisations. The first sampled point is $x_1 = (0.22, 0.41)$.

3.1. History-dependent rejection model

First, we demonstrate the history-dependent rejection model with self-interaction defined through the convex hull coverage, where the points outside the convex hull of the current point set are favoured according to the reweighting probability (5). Second, we demonstrate the rejection model with recurrence self-interaction by using the reweighting probability (8), which takes the number of generated points near the suggested point into account.

3.1.1. Coverage self-interaction

The purpose of this example is to illustrate self-interaction caused by the parameter ρ in the history-dependent rejection model with convex hull coverage (6). We fix the parameter $\sigma^2 = 0.3$ of the truncated Gaussian kernel (3) and vary the parameter ρ :

- *Model a*, $\rho = 1$ (random walk without self-interaction).
- *Model b*, $\rho = 0.1$ (fast coverage), which accepts points inside the convex hull of previous points with low probability.
- *Model c*, $\rho = 0.5$ (mild coverage), which accepts points inside the convex hull of previous points with mild probability.

We simulate 19 realisations of the random walk model *a*, since it here represents a reference model, and five realisations of Model *b* and Model *c*. One of the simulated realisations of each model can be seen in Figure 2. The polygons in the figure illustrate the convex hull coverage related to the first 10 points of the process. One cannot detect much difference between the random walk model *a* and mild coverage model *c*, but the fast expansion of Model *b* can perhaps be seen: there are early (dark) points near the edges.

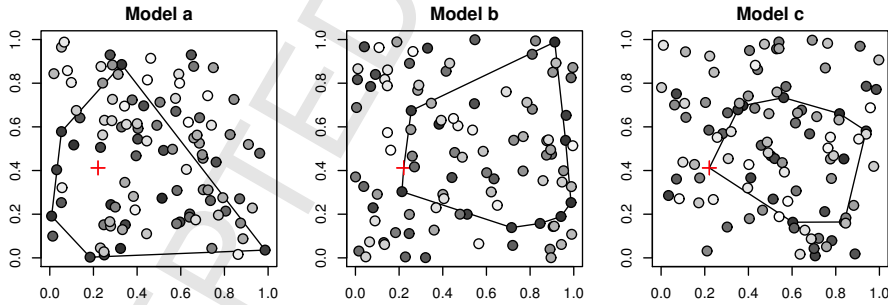


Figure 2: Simulated patterns of the three history-dependent models with coverage self-interaction. The colour of the points denotes their order (from dark to light) and the first fixation is marked with a cross. The polygon indicates the convex hull of the first 10 points.

However, point patterns only tell us about the spatial nature of the point process. Since we are mainly interested in the sequential (time order) aspect, we use four different functional summary statistics: ball union coverage (with radius 0.1), convex hull coverage, scanpath length and cumulative recurrence. (The two latter ones are explained below). The results related to these summaries can be found in Figure 3. The ball union coverage does not distinguish between the three models, but the convex hull coverage reveals that the coverage

of Model b increases faster than for the other two models. Accordingly, Model b makes longer jumps on average than Model a or c , as can be seen from the scanpath length which measures the length of the sample path cumulatively.

The recurrence function used in the reweighting probability (8) calculates the numbers of points near the current point excluding the previous point, and the cumulative version sums all these numbers together. Now we see that Model b avoids the locations nearby other points when compared to the random walk. This may be due to the fact that the fast coverage process b penalizes slow coverage and hence increases the drift of points near the edges.

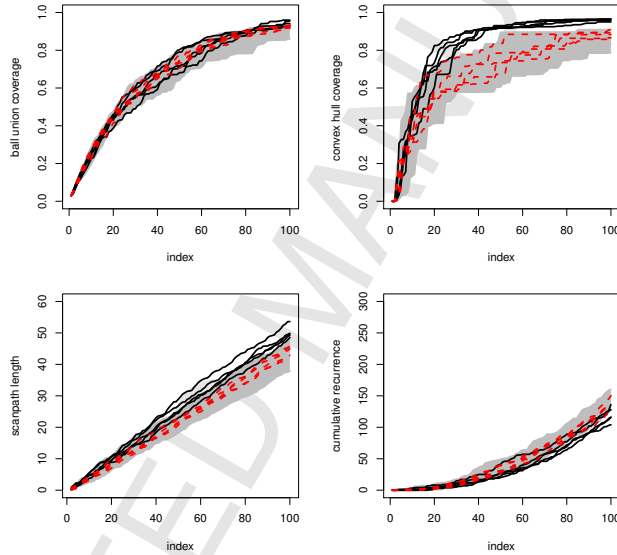


Figure 3: Ball union coverage with radius 0.1 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 0.1 (bottom right) for the two models (five realisations of each): The fast coverage model b is marked with black solid lines and the mild coverage model c is marked with red dashed lines. The grey area represents the envelopes estimated from 19 realisations of the random walk model a used as the reference model.

3.1.2. Recurrence self-interaction

Here we illustrate self-interaction in the history-dependent rejection model with recurrence (9). We again fix the parameter σ^2 of the truncated Gaussian kernel (3) to 0.3 and vary the self-interaction parameter θ :

- Model d , $\theta = 0.5$ (random walk without self-interaction).
- Model e , $\theta = 0.1$ (low recurrence), which favours points in the non-visited areas rather than the points in the areas nearby the previous points.

- *Model f*, $\theta = 0.9$ (high recurrence), which accepts points nearby the previous points with high probability.

We again simulate 19 realisations of the random walk model *d* as well as five realisations of Model *e* and Model *f*. In Figure 4 we can see that, compared with the random walk model *d*, the realisation of the low recurrence model *e* indicates a tendency towards higher regularity, and the realisation of the high recurrence model *f* is clearly more clustered. Note also that Model *e* seems to cover the whole area quite fast compared with the other two models.

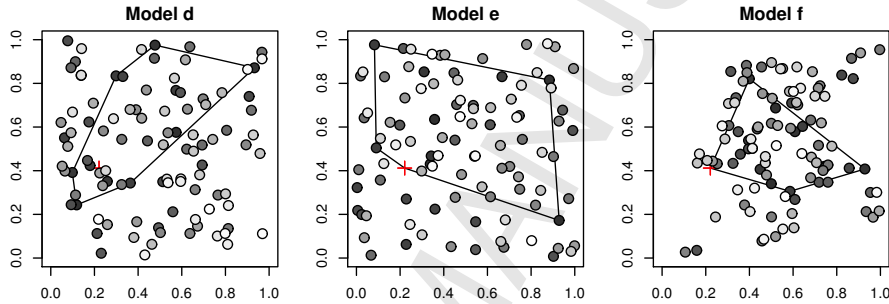


Figure 4: Simulated patterns of the three history-dependent models with recurrence self-interaction. The colour of the points denotes their order (from dark to light) and the first fixation is marked with a cross. The polygon indicates the convex hull of the first 10 points.

The results of the functional summary statistics are depicted in Figure 5. The ball union coverage describes clustering of the points, hence for the high recurrence model *e* the ball union coverage curves locate lower than for the low recurrence model *f*, which covers the whole window with 100 points. The convex hull coverage curves reveal an effect similar to the ball union coverage: the low recurrence model *e* almost fills the whole unit square, whereas the high recurrence model *f* only fills around 60 % of the area.

There is not much difference between the processes when comparing the scan-path lengths for the first 40 points, but after that the high recurrence model *f* makes shorter jumps on average compared with the other two models. However, the cumulative recurrence function clearly reveals that the two models differ from the random walk model: the low recurrence model *e* avoids areas close to the previous included points, while the high recurrence model *f* favours areas near the previous included points.

3.2. History-adapted model with convex hull coverage self-interaction

In this example we fix the kernel width parameter $\tau = 0.3$ and pay attention to the effect of the parameter κ of the history-adapted model (12)-(13). The

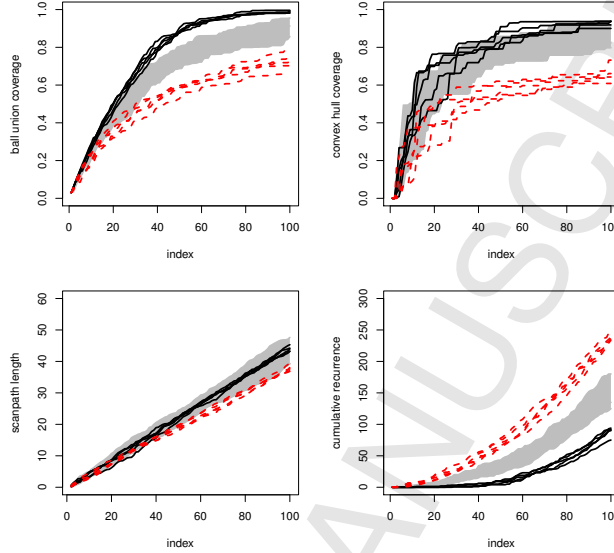


Figure 5: Ball union coverage with radius 0.1 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 0.1 (bottom right) for the two models (five realisations of each): The low recurrence model e is marked with black solid line and the high recurrence model f is marked with red dashed line. The grey area represent the envelopes estimated from 19 realisations of the random walk model d .

parameter κ controls the speed of decay as a function of coverage. We again define three history-adapted models:

- *Model g*, $\kappa = 0$ (random walk, the kernel does not change in time).
- *Model h*, $\kappa = 2$ (mild clustering), which means that the process is allowed to take long jumps at the beginning, but eventually starts to cluster.
- *Model i*, $\kappa = 4$ (fast clustering), which starts to cluster rather quickly when the coverage increases.

The kernel function (12) here uses the convex hull coverage, which means that in (13) $S(\vec{x}_k)$ is the area of the convex hull coverage generated by the points (x_1, \dots, x_k) . Now the history-adapted model works in such a way that at first the kernel width parameter τ is dominating and the process can make long jumps, but when the area of the convex hull of points approaches the size of the window, the parameter κ starts to affect and produces clustering.

In Figures 6 and 7, it can be seen that the convex hull of the first 10 points is of the same size for all models: the speed of coverage seems to be similar for all processes at the beginning. The spatial structure of the mild clustering model

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

h and the random walk model g are quite similar, but the points of the fast clustering model i are clearly more clustered than the points of the other two models, and there are only a few points in the upper half of the unit square.

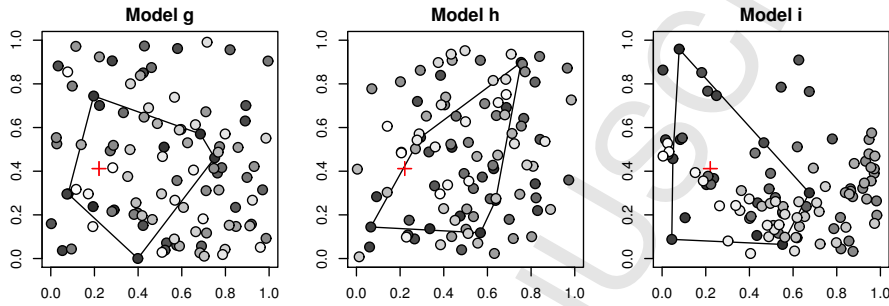


Figure 6: Simulated patterns of the three history-adapted models. The colour of the points denotes their order (from dark to light) and the first fixation is marked with a cross. The polygon indicates the convex hull of the first 10 points.

The functional summary statistics are plotted in Figure 7. The fast clustering model i covers the area similarly to the other two models at the beginning, but after about 50 points it starts to cluster, which can be seen as a decline of the ball union coverage summaries. The convex hull coverage does not reveal much difference between the models, and all the models are able to cover at least 60 % of the window. This is due to the wide kernel ($\tau = 0.3$) which allows the processes to make long jumps at the beginning.

The summary statistic that shows the clearest difference between the three models is the scanpath length. While the jumps in the random walk model have time invariant transitions, the clustering models h and i start to take shorter jumps at some point, which is indicated by the decline in scanpath curves. In addition, the cumulative recurrence function shows that the fast clustering model i gathers points around the previous ones. To conclude, the effect of the decay parameter κ seems to fasten the clustering as a function of coverage.

4. A case study: Black Bow by Wassily Kandinsky

We apply the developed modelling to experimental eye movement data related to arts in order to study self-interaction. The participants of the art experiment were inspecting six pictures of paintings and their eye movements were recorded. The stimulus picture was shown on the screen with a 1024×768 resolution and the eye movements were measured by the SMI iView XTM Hi-Speed eye tracker with temporal resolution of 500 Hz. The distance between a participant's head and the screen was about 85cm, and a forehead rest was used in order to prevent

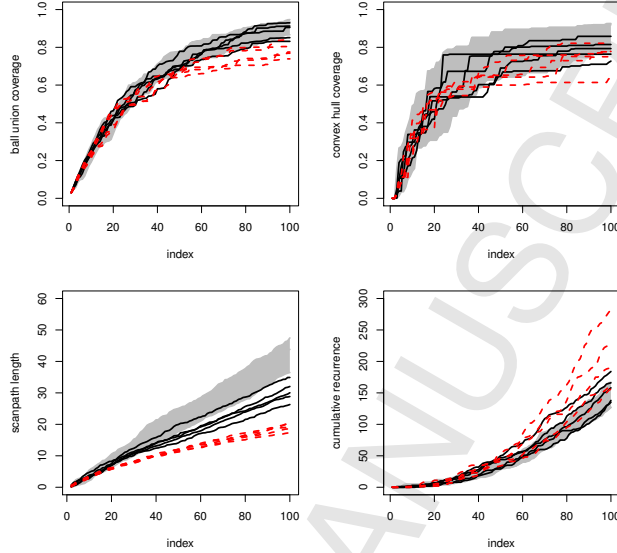


Figure 7: Ball union coverage with radius 0.1 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 0.1 (bottom right) for the two models (five realisations of each): The mild clustering model h is marked with black solid lines and the fast clustering model i is marked with red dashed lines. The grey area represents the envelopes estimated from 19 realisations of the random walk model g used as the reference model.

unintentional head movements. Each stimulus painting was shown for three minutes. The participants were also asked to describe the moods of the painting and their voice was recorded, but this information is not used here.

We will focus on one painting used in the experiment, called *Black Bow* (1912) by Wassily Kandinsky shown in Figure 1 (source of the painting: DÜCHTING (1991)). We will fit four versions of the history-dependent model with recurrence self-interaction to the eye movement data of one subject. The goodness-of-fit of the model is checked using the four functional summary statistics mentioned earlier, and the best fitting model is compared with the other subjects' data in order to conclude whether the same model fits well for all participants.

4.1. Fitting the history-dependent rejection model with recurrence self-interaction

We choose one subject of which eye movements are analysed and modelled here. Because of the long inspection period (three minutes), we decided to use only 100 first fixations of the sequence corresponding to a 35 second time-interval, as shown in Figure 1. According to the two-stage model (see e.g. Locher et al.,

1996) the overall impression of the scene is obtained during the first few fixations, and then the focus turns to the presumably interesting features. In addition, the gaze has a tendency to return to the interesting parts of the scene. Our aim is to find out whether we can find this sort of behaviour, i.e. if the process is of self-interacting type and if we can catch it with our rejection model.

We first investigate the variation of the four functional summary statistics related to this particular data from Kadinsky's painting. The ball union coverage with radius of 35 pixels, convex hull coverage, scanpath length and cumulative recurrence (radius 50) of the 20 subjects of the experiment are presented in Figure 13, as a dark solid curve for the subject under study, and as grey curves for the other participants. It can be noticed that the first 100 fixations do not cover the whole painting (the ball union covers around 30 % and the convex hull around 40 % of the target). It is typical of the eye movements that the edges of the painting are avoided and that is why the coverage hardly ever reaches the whole scene.

Next, we estimate the heterogeneity term $\alpha(x)$ for the target painting. In this case, we utilize the empirical saliency map estimated as the intensity of fixation patterns of all the 20 subjects excluding the one under study (a total of 9366 fixations is used for the intensity estimation). Problems associated with the estimation of $\alpha(x)$ are considered in Discussion. For technical reasons $\alpha(x)$ is scaled to have values in $[0, 1]$. The scaled saliency map together with the fixations of the subject under study can be seen in Figure 8. This particular subject paid most attention to the areas with high intensity, but the gaze stayed in some low intensity areas also.

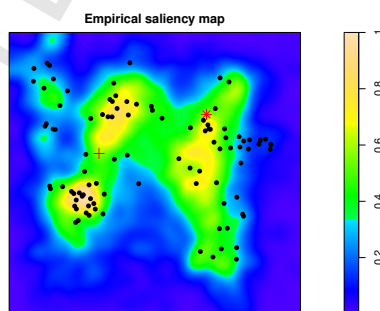


Figure 8: Empirical saliency map estimated from the fixations of all subjects excluding the one under study. The points indicate the first 100 fixation points of the particular subject under study. The first fixation is marked with a cross and the second with a star.

In what follows, we fit four different rejection models consisting of three components: heterogeneity (H), contextuality (C) and self-interaction (S). The

self-interaction is assumed to be caused by the delayed recurrence function $S_R(\vec{x}_k, x)$, $k = 2, \dots, n - 1$ defined in (7), where the radius $r = 50$ pixels is used. Therefore, we condition by the first two fixations, x_1 and x_2 . All four models are submodels of (9), see Table 1.

Model	Components	Parameter values
1	H	$\theta = \frac{1}{2}, \sigma^2 \text{ large}^*$
2	H, C	$\theta = \frac{1}{2}, \sigma^2 > 0$
3	H, S	$0 \leq \theta \leq 1, \sigma^2 \text{ large}^*$
4	H, C, S	$0 \leq \theta \leq 1, \sigma^2 > 0$

Table 1: Components of the four rejection models. (*The value of σ^2 should be chosen to be large enough such that the kernel is flat in the specified window.)

Model 1 includes heterogeneity and is a binomial process in a heterogeneous environment. When the Markovian kernel function is added (Model 2) the process is a random walk with Markovian property. Model 3 is a self-interacting process in a heterogeneous media without the contextuality effect and Model 4 contains both the Markov kernel and the self-interaction term.

The log-likelihood function for Model 4 is now

$$\begin{aligned}
l(\sigma^2, \theta) &= \sum_{k=2}^{n-1} \log(\alpha(x_{k+1})) - \frac{1}{2\sigma^2} \sum_{k=2}^{n-1} \|x_k - x_{k+1}\|^2 \\
&+ \log(1 - \theta) \sum_{k=2}^{n-1} \mathbf{1}_{\{S_R(\vec{x}_k, x_{k+1})=0\}}(x_{k+1}) + \log(\theta) \sum_{k=2}^{n-1} \mathbf{1}_{\{S_R(\vec{x}_k, x_{k+1}) \geq 1\}}(x_{k+1}) \\
&- \sum_{k=2}^{n-1} \log \int_W \alpha(u) e^{-\frac{1}{2\sigma^2} \|x_k - u\|^2} ((1 - \theta) \mathbf{1}_{\{S_R(\vec{x}_k, u)=0\}} + \theta \mathbf{1}_{\{S_R(\vec{x}_k, u) \geq 1\}}) du.
\end{aligned} \tag{16}$$

The likelihood for Model 1 is just the first term on the right hand side of equation (16). For Model 2, the likelihood is obtained choosing $\theta = \frac{1}{2}$ in (16) and for Model 3 choosing σ^2 to be large (i.e. large enough such that the kernel is flat in the specified window). The parameters are estimated using the coordinate ascent algorithm. We used numerical integration for computing the normalization term of the log-likelihood (16), last row, and a grid of (60, 80, ..., 400) and (0.05, 0.10, ..., 0.95) for maximizing the likelihood.

Model 1 includes only the empirical saliency map and we do not have to estimate any parameters. For Model 2 we get $\hat{\sigma} = 180$, and for Model 3 $\hat{\theta} = 0.75$. The parameter estimates for Model 4 are $\hat{\sigma} = 180, \hat{\theta} = 0.70$. Note that for a random walk model we should have $\theta = 0.50$; hence there are recurrence features involved in this data.

4.1.1. Model comparisons

Assessing the goodness-of-fit of the models is here done by estimating the four summary statistics mentioned earlier from the data and from 99 simulated realisations of the fitted models. When simulating the model, we condition on the observed values of the first two fixations in order to reduce variation right at the beginning of the process. One simulated realisation of each model can be seen in Figure 9, and the summary statistics estimated from the data with pointwise envelopes estimated from the simulations in Figures 10 – 13. Note that the ball union and convex hull coverages are here presented with respect to the size of the window, hence they obtained values in $[0, 1]$.

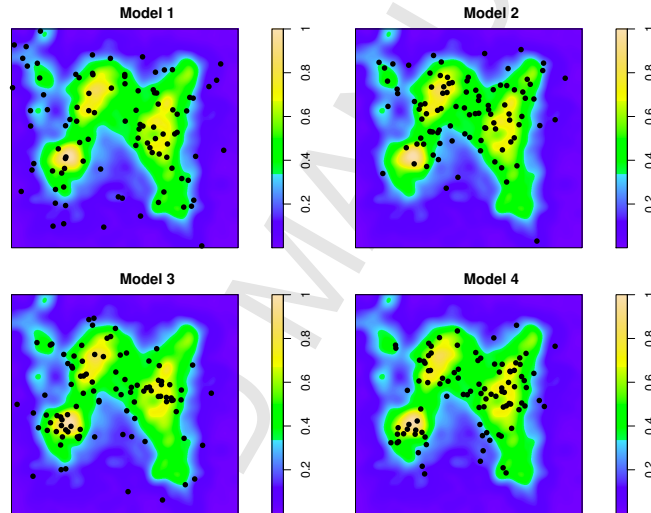


Figure 9: Simulated realisations of the four models fitted to the eye movement data of the particular subject under study, overlaid with the empirical saliency map based on data from the other subjects. The first two points are fixed and marked with a cross and a star, respectively.

For Model 1 all summary statistics show poor fit (Figure 10). Compared with the data this model covers the target area too fast, takes too long jumps according to the scanpath length, and goes to areas with too few points according to the cumulative recurrence function. As a conclusion, the heterogeneity component alone does not describe the data set well enough even though the spatial heterogeneity is followed rather well (see Figure 9 upper left).

Model 2 seems to perform slightly better than Model 1. The summaries estimated from the data set stay inside the simulated pointwise envelopes, except the ball union coverage and cumulative recurrence function after the first 70 points (Figure 11). These findings indicate that data begin to cluster at the end of the inspection period, but the model does not carry that effect.

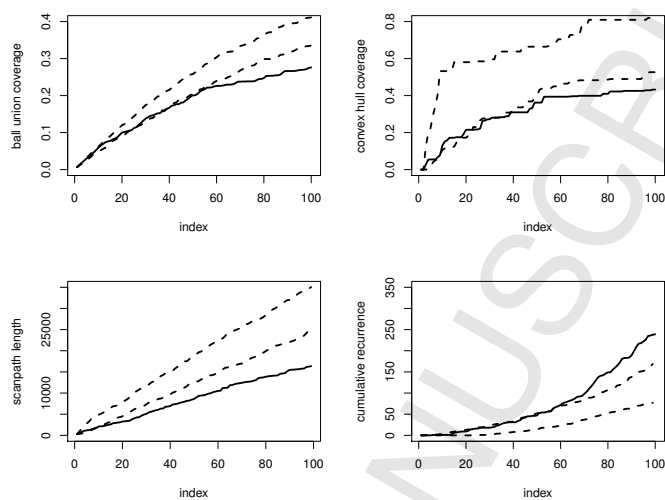


Figure 10: Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (solid line). Dashed lines represent pointwise envelopes estimated from 99 simulations of Model 1.

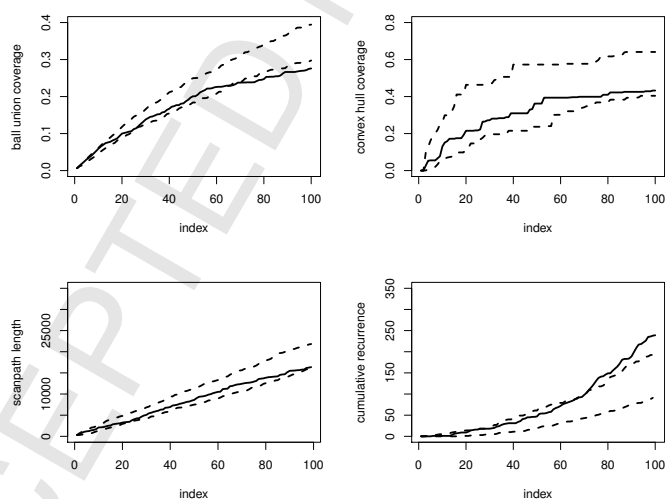


Figure 11: Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (solid line). Dashed lines represent pointwise envelopes estimated from 99 simulations of Model 2.

Model 3 includes heterogeneity and self-interaction, but not contextuality, which is related to the length of the jumps the process makes. As a result, this model

seems to jump too much compared with data, since the estimated scanpath length summary is at odds with the simulated pointwise envelopes (Figure 12). The marginal spatial structure looks slightly more clustered than Model 1 and Model 2 predict (Figure 9).

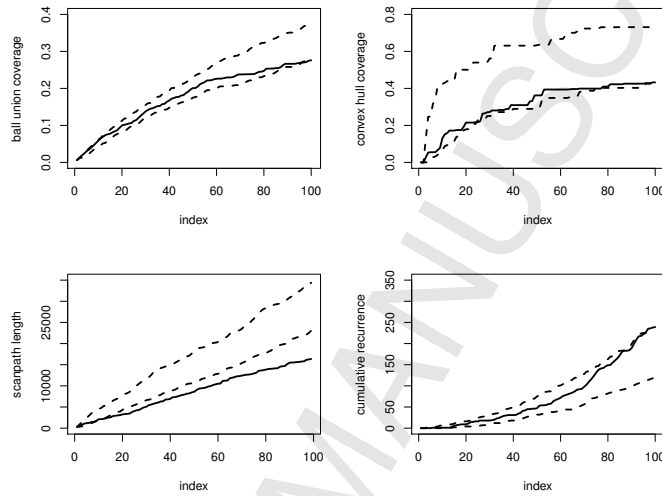


Figure 12: Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (solid line). Dashed lines represent pointwise envelopes estimated from 99 simulations of Model 3.

Model 4 includes all the three effects and seems to be in good agreement with data: all four summary statistics estimated for the subject under study stay within the simulated pointwise envelopes, see Figure 13. It seems that this model is able to catch the nature of this eye movement process fairly well. The estimated parameter value $\hat{\theta} = 0.75$ indicates that the locations nearby the previous points (excluding the most recent point) are favoured, which is a cause of spatial clustering. We conclude that the random walk model does not seem to be a good model for these data, but there is self-interaction due to the recurrence involved: the eye movement process seems to favour areas close to previous fixations.

4.1.2. Population level comparison

We have been able to describe the variation in the eye movement sequence of an individual by using the rejection model with recurrence-based weighting. In order to investigate the generality of the suggested models, we make comparisons at the population level using all 20 subjects. As can be seen in Figure 13, the envelopes of Model 4 seem to cover the convex hull and ball union coverage

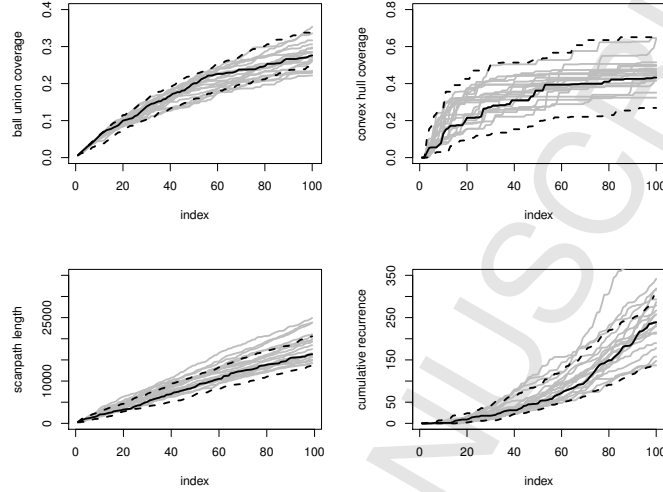


Figure 13: Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (black solid line) and for all other subjects (grey solid lines). Dashed lines represent pointwise envelopes estimated from 99 simulations of Model 4.

curves of the subjects rather well. The scanpath length does not cover the curves as well: there are subjects whose gaze makes longer jumps at the beginning of the eye movement process than the fitted model predicts. Furthermore, the envelopes of the cumulative recurrence function cover almost all the curves, but there is one exceptional subject, whose fixations are strongly clustered after the 70th fixation.

In order to further describe the variation in the data set related to self-interaction, we fitted Model 4 for the first 100 fixations of each subject separately. The estimates of the parameters σ and θ can be seen in Table 2. The 90 % bootstrap confidence intervals for the parameter estimates are calculated from 20 realisations of the fitted model. When the parameter σ is large, the model allows jumps over the observation window, and then the self-interaction parameter θ dominates. Large θ indicates strong spatial clustering. When σ is small, one needs to take multiple jumps in order to cross the whole target window.

In Figure 13, we have plotted the estimated summary statistics for all subjects together with the pointwise envelopes based on the model fitted to subject 5. For subjects 13, 18 and 20, the scanpath length curves clearly exceed the envelopes of the Model 4 fitted for subject 5. For each of them, the fitted value of the parameter σ in Model 4 is over 280. This indicates that the process is allowed to make long jumps resulting in longer scanpaths. However, for these subjects the recurrence parameter θ varies from 0.70 to 0.80 and does not differ much from the recurrence parameter of subject 5.

Subject id	$\hat{\sigma}$	90 % CI for $\hat{\sigma}$		$\hat{\theta}$	90 % CI for $\hat{\theta}$	
1	240	200	280	0.75	0.70	0.85
2	180	160	200	0.55	0.45	0.65
3	160	140	180	0.65	0.60	0.75
4	160	140	180	0.80	0.75	0.85
5	180	140	200	0.70	0.65	0.85
6	180	140	200	0.85	0.80	0.95
7	160	140	180	0.70	0.65	0.80
8	240	200	280	0.70	0.65	0.80
9	160	140	180	0.65	0.60	0.75
10	160	140	180	0.65	0.60	0.75
11	200	180	220	0.70	0.65	0.80
12	260	240	320	0.85	0.80	0.90
13	280	220	340	0.70	0.60	0.85
14	220	200	240	0.70	0.60	0.75
15	220	180	260	0.55	0.45	0.60
16	200	160	220	0.70	0.65	0.80
17	160	140	180	0.70	0.60	0.80
18	340	220	420	0.80	0.75	0.90
19	140	120	160	0.70	0.65	0.80
20	280	200	340	0.80	0.70	0.85

Table 2: Estimated parameters $\hat{\sigma}$ and $\hat{\theta}$ of Model 4 with their confidence intervals for all subjects. The subject under closer study is number 5 (bold).

For subject 6, the cumulative recurrence curve is way above the envelopes, and the estimated recurrence parameter is 0.85 indicating strong clustering. The other subjects, whose cumulative recurrence curve is outside the envelopes, are 4, 12, 18 and 20, and for all of them the recurrence parameter is over 0.80. As a conclusion, there is some variation related to the clustering effect of the points between these subjects. However, for each subject the estimated recurrence parameter differs from 0.5 meaning that the random walk is not a suitable model.

5. Discussion

In this paper, we develop advanced data-analytical tools for extracting information from eye movement sequences needed in various areas of application utilizing eye tracking (see e.g. Rayner, 2009). Our objective is to create simple but flexible dynamic stochastic models by employing mechanisms which use the whole history of the sequence in each gaze jump in order to capture features of learning during the experiment.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Heterogeneity of the scene, contextuality of subsequent fixations, and self-interaction of eye movements are elements that affect the eye movement process. We present a sequential spatial point process approach which includes these three effects, the self-interaction being new in this context and is interpreted as a learning effect. This leads to what in probability theory is called self-interacting processes, which are generalizations of random walks in heterogeneous media. Although self-interacting random walks are well established in mathematics, physics and animal ecology, our reasoning here is slightly different. We study how the process evolves at an early stage of an eye movement sequence whilst, e.g. in mathematics, the long term behaviour is of interest. Such processes are analytically difficult, even intractable, but their simulation is basically straightforward.

After having constructed a new model, we need model fitting (estimation), evaluation of goodness-of-fit (model criticism) and simulation algorithms for various inferential purposes. Here, we suggest a likelihood approach for the new processes which is used in parameter estimation. It is not possible to obtain analytical results or use asymptotical reasoning. Instead, we compute the likelihood using simulation which allows us to make exact inference in the sense that it does not depend on the size of data and which includes a boundary effect correction. In doing this, we enlarge the applicability of spatial statistics and the likelihood inference to a new area of applications. The processes can be used to make inference on the structure of data, including self-interaction, and to deduce uncertainties in conventional and new functional data summaries such as scanpath length and recurrence function.

In this paper, we are interested in the dynamics of the eye movement process. The main question is whether there is self-interaction present in a given eye movement sequence and whether we can detect it using our new modelling. We focus on the beginning of an eye movement process, since, according to the two-stage model by Locher and colleagues (see e.g. Locher, 2006), the gist of the scene is established during the early fixations. Our history-dependent rejection model is, in fact, able to observe self-interaction in these particular data, although there is large within-subject variation.

Our models utilize stochastic geometry in creating self-interaction caused either by coverage (how much of the scene is covered and how fast) or by recurrence (how much the process favours points nearby the previous points), both of which have justifications in eye movement literature. Functional summary statistics are needed for checking the goodness-of-fit of a fitted model, as well as for describing the structural components of the sequence. We use four summary statistics: convex hull coverage, ball union coverage, scanpath length, and cumulative recurrence. Several summary statistics are needed since none of these four was able to alone distinguish between the models in our simulation study. We found that the rejection model with convex hull coverage can be separated from the random walk by the scanpath length, whereas the rejection model with

recurrence can be distinguished from random walk with the coverage measures. The scanpath length and cumulative recurrence are needed with the history-adapted model for defining the speed of spatial clustering.

We have illustrated two tractable process constructions for self-interaction, namely, history-based independent thinning and history-dependent transitions. These constructions are very different, and their use depends on the problem and the data set. The two developed models are rather simple but can easily be extended. Other constructions are also possible, such as the heterogeneous mixture model, where the Markov kernel $K(\vec{x}_k, x)$ is replaced by

$$p(\vec{x}_k) K_1(x_k, x) + (1 - p(\vec{x}_k)) K_2(x_k, x).$$

Here, K_1 and K_2 are two kernel functions where the choice $K_2(x_k, x)$ could be uniform in W or proportional to $\alpha(x)$, for example, and the mixture factor $p(\vec{x}_k)$ depends on the history of the sequence. Although simulation of such a model is straightforward, the associated inference is computationally demanding.

We have restricted our approach to sequential spatial point processes, mainly due to their tractability. However, this approach is a bridge to spatio-temporal models that would take fixation durations into account. For a separable spatio-temporal model, the spatial effect and time dynamics are multiplicative in the likelihood. A sequential spatial point process model can be used as a building block: if an order-dependent spatial model is available, the inclusion of time dynamics is straightforward, because inference on the ordered spatial aspect and fixation durations can be performed independently. If a preferred summary statistic contains information on the fixation durations, then a spatio-temporal model should be used instead of the sequential point process.

The estimation of the heterogeneity component $\alpha(x)$ is problematic. In the second order analysis of point patterns the first and second order components are not estimable from one observed point pattern without further information. Diggle et al. (2007) suggest two alternatives, which are the use of a parametric model for the intensity (heterogeneity) or, alternatively, the utilization of replications for the intensity estimation. In the sequential context the situation is similar. In our experiment sequences measured from several participants are available and are independent of the particular sequence under study. When using this information in the estimation of the heterogeneity component, the problem is that also these auxiliary control sequences are serially correlated leading to extra clustering at the sequence level. We assume that this effect is not as serious as in the intensity estimation from the case data only. When using auxiliary sequences, we have assumed that these sequences contain information which origins mainly from the target common to all participants and to the case under study and measures the wanted heterogeneity.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

An improvement would be to sample fixations from each of the auxiliary fixation sequences instead of using all the fixations as we did here. Merging these sampled fixations gives a point pattern which is then used in the estimation of $\alpha(x)$ using the kernel method. This will further reduce the effect of serial correlation. An alternative improvement is based on the case sequence under study by using the fitted model (containing both contextuality and self-interaction and a prefixed $\alpha(x)$) to compute the inverse of the transition probability for each fixation. These weights can then be used in the estimation of $\alpha(x)$ by the weighted kernel method. The procedure can be iterated. The estimation of $\alpha(x)$ is discussed in Barthelmé et al. (2013); Engbert et al. (2015).

Another issue concerns the parameter estimation. Here, we suggest the discretized coordinate ascent algorithm for maximum likelihood using forward simulation. This early experimenting shows that by this method it is possible to separate the effect of self-interaction and present confidence intervals for parameter estimates and confidence envelopes for chosen summary statistics. We know that approximative inference, being computationally much faster, is also a possibility and would be very important in a methodological toolbox.

Acknowledgements

The authors would like to thank the three anonymous reviewers for their valuable comments, which helped in improving the manuscript. As a consequence, we have paid attention to the estimation of the heterogeneity component of the models. The authors are also grateful to Aila Särkkä for the useful comments and suggestions, to Tuomas Rajala for his help with the computational issues, and to Professor Pertti Saariluoma, Sari Kuuva, María Álvarez Gil, Jarkko Hautala, and Tuomo Kujala for providing the data set.

The second author has been financially supported by the Finnish Doctoral Programme in Stochastic and Statistics and by the Academy of Finland (Project number 275929).

Baddeley, A., Rubak, E., Turner, R., 2015. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London.

Baddeley, A. J., Diggle, P. J., Hardegen, A., Lawrence, T., Milne, R. K., Nair, G., 2014. On test of spatial pattern based on simulated envelopes. *Ecological Monographs* 84 (3), 477–489.

Baddeley, A. J., van Lieshout, M. N. M., 1995. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* 47 (4), 601–619.

- 1
2
3
4
5
6
7
8
9 Barthelmé, S., Trukenbrod, H., Engbert, R., Wichmann, F., 2013. Modelling
10 fixation locations using spatial point processes. *Journal of Vision* 13 (12),
11 1–34.
- 12
13 Diggle, P. J., 2013. *Statistical Analysis of Spatial and Spatio-Temporal Point*
14 *Patterns*. CRC Press, Boca Raton.
- 15
16 Diggle, P. J., Gómez-Rubio, V., Brown, P. E., Chetwynd, A. G., Gooding, S.,
17 2007. Second-order analysis of inhomogeneous spatial point processes using
18 case-control data. *Biometrics* 63 (2), 550–557.
- 19
20 Diggle, P. J., Kaimi, I., Abellana, R., 2010. Partial-likelihood analysis of spatio-
21 temporal point-process data. *Biometrics* 66 (2), 347–354.
- 22
23 Düchting, H., 1991. *Wassily Kadinsky 1866-1944: Maalaustaiteen vallankumous*
24 (Transl. by T. Braun). Köln. Benedikt Taschen, p. 56.
- 25
26 Efron, B., Tibshirani, R. J., 1994. *An introduction to the Bootstrap*. Chapman
27 and Hall/CRC Press.
- 28
29 Engbert, R., Trukenbrod, H. A., Barthelmé, S., Wichmann, F. A., 2015. Spatial
30 statistics and attentional dynamics in scene viewing. *Journal of Vision* 15 (1),
31 1–17.
- 32
33 Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates
34 of the variance of united kingdom inflation. *Econometrica: Journal of the*
35 *Econometric Society* 50 (4), 987–1007.
- 36
37 Geyer, C. J., 1991. Markov chain Monte Carlo maximum likelihood. *Interface*
38 *Foundation of North America*. Retrieved from the University of Minnesota
39 Digital Conservancy, <http://purl.umn.edu/58440>.
- 40
41 Grabarnik, P., Myllymäki, M., Stoyan, D., 2011. Correct testing of mark in-
42 dependence for marked point patterns. *Ecological Modelling* 222 (23), 3888–
43 3894.
- 44
45 Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. *Statistical analysis and*
46 *modelling of spatial point patterns*. Chichester: Wiley.
- 47
48 Itti, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert
49 shifts of visual attention. *Vision Research* 40, 1489–1506.
- 50
51 Kümmerer, M., Wallis, T., Bethge, M., 2014. How close are we to understanding
52 image-based saliency? arXiv preprint arXiv:1409.7686.
- 53
54 van Lieshout, M. N. M., 2006a. Campbell and moment measures for finite se-
55 quential spatial processes. *Report-Probability, networks and algorithms* (1),
56 1–10.
- 57
58 van Lieshout, M. N. M., 2006b. Maximum likelihood estimation for random
59 sequential adsorption. *Advances in Applied Probability (SGSA)* 38, 889–898.

- 1
2
3
4
5
6
7
8
9 van Lieshout, M. N. M., 2009. Sequential spatial processes for image analysis.
10 In: *Stereology and Image Analysis. Ecs10. Proceedings of the 10th European*
11 *Congress of ISS. Bologna, Italy.*
- 12
13 Locher, P. J., 2006. The usefulness of eye movement recordings to subject an
14 aesthetic episode with visual art to empirical scrutiny. *Psychological Science*
15 48, 106–114.
- 16
17 Locher, P. J., Gray, S., Nodine, C. F., 1996. The structural framework of picto-
18 rial balance. *Perception* 25 (12), 1419–1436.
- 19
20 Locher, P. J., Krupinski, E. A., Mello-Thoms, C., Nodine, C. F., 2007. Visual
21 interest in pictorial art during an aesthetic experience. *Spatial Vision* 21,
22 55–77.
- 23
24 Rayner, K., 2009. Eye movements and attention in reading, scene perception,
25 and visual search. *The Quarterly Journal of Experimental Psychology* 62,
26 1457–1506.
- 27
28 Ripley, B. D., 1987. *Stochastic simulation.* Wiley.
- 29
30 Tatler, B. W., Baddeley, R. J., Vincent, B. T., 2006. The long and the short of
31 it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision*
32 *Research* 46 (12), 1857–1862.
- 33
34 Ylitalo, A.-K., Särkkä, A., Guttorp, P., in press. What we look at in paintings:
35 A comparison between experienced and inexperienced art viewers. *Annals of*
36 *Applied Statistics.*
- 37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65