UNIVERSITY OF JYVÄSKYLÄ
DEPARTMENT OF MATHEMATICS
AND STATISTICS

REPORT 160

UNIVERSITÄT JYVÄSKYLÄ
INSTITUT FÜR MATHEMATIK
UND STATISTIK

BERICHT 160

STATISTICAL INFERENCE FOR EYE MOVEMENT SEQUENCES USING SPATIAL AND SPATIO-TEMPORAL POINT PROCESSES

ANNA-KAISA YLITALO

To be presented, with permission of the Faculty of Mathematics and Science
of the University of Jyväskylä, for public criticism in Auditorium S212
on June 1st, 2017, at 12 o'clock noon.

JYVÄSKYLÄ
2017

Editor: Pekka Koskela
Department of Mathematics and Statistics
P.O. Box 35 (MaD)
FI–40014 University of Jyväskylä
Finland

ISBN 978-951-39-7063-5 (print)
ISBN 978-951-39-7064-2 (pdf)
ISSN 1457-8905

University Printing House
Jyväskylä 2017

Abstract

Eye tracking is a widely used method for recording eye movements, which are important indicators of ongoing cognitive processes during the viewing of a target stimulus. Despite the variety of applications, the analyses of eye movement data have been lacking of methods that could take both the spatial and temporal information into account. So far, most of the analyses are based on strongly aggregated measures, because eye movement data are considered to be complex due to their richness and large variation between and within the individuals. Therefore, the eye movement methodology needs new statistical tools in order to take full advantage of the data.

This dissertation is among the first studies to employ point process statistics for eye movement data in order to understand its spatial nature together with the temporal dynamics. Here, we consider eye movements as a realisation of a spatio-temporal point process. The emphasis is in statistical inference on eye movements using existing point process statistics along with the new methods and models introduced in this work. Our aim is to get understanding of eye movements as a temporally evolving process in space. This objective is achieved in four steps: First, we apply the second-order characteristics of point processes to describe features of the process. Second, we develop new functional summary statistics in order to evaluate the temporal nature of the eye movements. Third, we use likelihood-based modelling to assess the uncertainty related to these data summaries. Fourth, the developed models are used both for group comparisons and for distinguishing components in an eye movement sequence.

The empirical results of this dissertation give new information on visual processing of paintings. We find evidence that the viewing process of one subject changes during the inspection of the painting being an indication of learning. The behaviour of this learning effect, however, varies between the individuals. We also study differences between novices and non-novices in art viewing by comparing where they look at and for how long the gaze typically stops. The latter distinguishes the two groups, whereas the former reveals minor differences that are not statistically significant. Altogether, we hope that our results encourage researchers to pay more attention to temporal dynamics in eye movement data, as well as to the inevitable variation in the individual level.

The spatio-temporal analysis of eye movements presented here is novel and covers a wide range of methods from functional summary statistics to the likelihood-based modelling. The methods and tools presented are applicable to other eye movement data collected in a free-viewing condition, but we believe that the developed models, being rather simple but flexible, could also be useful for the analysis of spatio-temporal sequences outside the field.

Acknowledgements

I owe my deepest gratitude to my primary supervisor, Professor Emeritus Antti Penttinen, who has been guiding me all the way from my master's thesis to the end of this PhD. Thank you for sharing your broad expertise, for keeping your door always open for my questions, for helping me with all kind of issues during my studies and for encouraging me to go abroad. I truly appreciate the time you have spent reading and commenting my work or having a cup of coffee with me. Not every supervisor is as committed to their job as you are. Thank you for having faith in me and this topic.

I would like to thank Associate Professor Matti Vihola for hopping in as my supervisor for the final part of my studies. Thank you for giving valuable comments on my thesis and organising all the practical things during the past months. To my secondary supervisor, Dr. Salme Kärkkäinen, I owe thanks for all the support she has given me along my studies, whether the issue has been an academic or a personal one. Thank you also for making my first conference trip so memorable.

I have had the privilege to collaborate with amazing researchers during this project. First, I would like to express my gratitude to Professor Aila Särkkä for welcoming me to Gothenburg without even knowing me, guiding me on my way from a student to a researcher, and being not only a collaborator but a friend as well. Second, I want to thank Professor Peter Guttorp for opening my eyes to the realisation that simple is beautiful when it comes to modelling. Thank you both for sharing your expertise with me. It has been insightful and fun.

I would like to thank Professor Pertti Saariluoma for introducing the problem of how to analyse eye movement data to my primary supervisor over a lunch. Without the eye tracking data you provided with Dr. Sari Kuuva at our disposal this thesis would not exist at all. Thank you, Sari, for the introduction to eye tracking in the early days of this project.

I would also like to thank Professor Marie-Colette van Lieshout and Professor Juha Heikkinen for reviewing this thesis.

This work was funded by the Jyväskylä Doctoral Program in Computational and Mathematical Sciences (COMAS), the Finnish Doctoral Programme in Stochastics and Statistics (FDPSS), and the Department of Mathematics and Statistics in University of Jyväskylä. I owe my gratitude also to Vilho, Yrjö and Kalle Väisälä foundation and University of Jyväskylä for the travel grants and to the Academy of Finland for funding me during the final stage of this project. All the conference trips, workshops and research visits gave me a bit more confidence and helped me to grow as a researcher. Not to mention the valuable comments I got about my work or the wonderful people I met during the travels.

I am lucky to have academic siblings with a background in point processes. Tuomas, thank you for sharing an office, helping me with computational issues, providing comments on my papers and spreading your sarcastic humour. Thank you, Mari, for keeping in contact and

asking how I'm doing. Your emails boosted my work flow.

To my girls in Sweden, Maud, Henrike, Mariana, I want to express my gratitude for all the encouragement and faith that you had in me. I wish I had a combination of Maud's confidence, Henrike's efficiency and Mariana's joy of life as my superpower. I hope we'll reunite many more times, regardless of how far from Gothenburg we end up to.

The Department of Mathematics and Statistics offered a good research environment for a PhD student. I would like to thank all my colleagues in statistics for your presence, support and encouraging words during my studies. It was a pleasure to have you as teachers and later as colleagues. Special thanks to you whose office I have bumped into for some peer support and to the people that I had so much fun with during the coffee breaks. Those interactions made my day.

I had a great opportunity to collaborate with Professor Erkki Huovinen and Dr. Marjaana Puurtinen in their music reading research during my studies. Now this collaboration has evolved to a consortium project funded by the Academy of Finland. I want to thank you both for including me in this project that has given me more than I ever expected. It has been challenging but also rewarding to work with different kinds of eye tracking data than the one in this thesis.

There are many more people that have been involved in my life during this bumpy road. Those with whom I have shared and office, studied, had lunch or coffee, talked about my research or escaped some locked rooms – thank you for the memorable moments. To my current colleagues at Musica; it has been a pleasure to get to know you all. I though of myself as an outlier among you, thanks for proving me wrong. I would also like to thank Tuulevi, Vesa and Sampo for their hospitality during the most exiting trip in my life. I cannot express well enough how grateful I am to you.

One group of friends has been a substantial part of my everyday life. A big thanks to DPoF for being continuously online and ready for a chat. Special thanks go to the Jyväskylä delegation for all the fun and relaxing moments that took my mind off work, and to Japan for sharing thoughts and feelings about PhD studies and academic life in general.

I am indebted to my parents for providing me a good premise for the path I chose. I would not be here today without the work ethic and perseverance that I absorbed from you. Thank you for the support and all the kilometres you have driven for me. To my brothers and their families I owe my gratitude for enriching my life.

My warmest gratitude goes to Ville for keeping me on track during this journey and making sure my everyday life stays somehow normal. I appreciate your dedication to my worries, no matter how irrational they were or what time of the day they came to my mind. Thank you for believing in me more than I ever did.

Jyväskylä, May 2017

Anna-Kaisa Ylitalo

List of original publications

This thesis consists of an introductory part and publications listed below.

I Ylitalo, A.-K., Penttinen, A., Särkkä, A. and Kuuva, S. Spatial point process statistics in understanding eye movement data. Submitted, 2017.

II Ylitalo, A.-K., Särkkä, A. and Guttorp, P. What we look at in paintings: A comparison between experienced and inexperienced art viewers. *The Annals of Applied Statistics*, 10(2): 549–574, 2016.

III Penttinen, A. and Ylitalo, A.-K. Deducing self-interaction in eye movement data using sequential point processes. *Spatial Statistics*, 17:1–21, 2016.

IV Ylitalo, A.-K. Detecting the long-memory effect in an eye movement sequence. Manuscript, 2017.

The author of this dissertation has actively taken part in the research and writing of the joint papers I–III. The research questions and statistical models have been formulated together with the co-authors. The author had the main responsibility in implementation of the models and in performing the simulation studies.

Contents

Chapter 1

Introduction

The present dissertation provides tools for statistical inference on eye movement data by employing spatial and spatio-temporal point process statistics and modelling. The novelty of this work is in considering the temporal evolution of an eye movement process without strong aggregation in space and in presenting model-based inference on the functional summary statistics describing the evolution. The main aims for this work are: to perform group comparisons, to assess the uncertainty of appropriate summary statistics and to detect long-term dependence in an observed eye movement sequence. For answering these questions, we use existing methods of point processes, introduce new summary statistics, develop sequential spatial and spatio-temporal point process models, do model fitting, and experiment with synthetic and real data. The results of this dissertation are expected to be useful for other applications measuring eye movements in a free-viewing condition, where the gaze is allowed to move freely in a two-dimensional space, but the models may be applied for other spatio-temporal data as well.

1.1 Tracking eye movements

Eye movements are reflections of cognitive processes in the brain and provide valuable information on where the attention is being directed to (Just and Carpenter, 1980). The movement of the gaze on a target is recorded by electronic devices, called eye trackers, which often locate below a computer screen. The eye tracking starts with a calibration phase, in which the eye tracker associates the location of the gaze with the target stimulus. Typically, the participant is asked to look at a set of calibration markers and the eye tracker records values from the eyes, such as pupil positions and corneal reflections, that are associated with the locations of the markers (Holmqvist et al., 2011). After the calibration, the eye tracker software is able to define the location of the gaze on the target stimulus with a sufficient accuracy and precision.

The eye tracking method has been applied in various psychological studies, such as reading, scene perception, visual search or music reading (see extensive reviews of Duchowski, 2002; Rayner, 1998, 2009), as well as in educational studies in recent years (see the review of Lai et al., 2013). Reasons for the on-going expansion of popularity of the method likely lie in the reduced prices and increased availability of eye trackers as well as in the development of user-friendly recording software produced by eye tracker manufacturers. Therefore, the eye tracking method is nowadays easy to apply without programming skills, and mobile equipments allow researchers to conduct experiments outside laboratories as well.

Despite the increasing use of eye tracking, the statistical methodology for analysing eye movement data is still rather limited. Simple eye movement analyses can be done by the eye tracker software that provides also comprehensive visualisations of eye movements. However, these analyses are useful mainly for describing the data by means of where people look at and for how long, since they are based on eye movement measures that are strongly aggregated over time and space. For text reading studies, the development of methodology and computational models for the eye movements are advancing rapidly (Rayner, 2009), but other areas of application come far behind.

Eye movements can be split into several kinds of events, of which the most reported one is called a *fixation* (e.g. Holmqvist et al., 2011). The fixation is a state, when the gaze is staying fairly still around a location over a time period. The duration of a fixation can vary from tens of milliseconds up to several seconds (Holmqvist et al., 2011). In practice, fixations are identified from raw eye movement data by algorithms, which use information on spatial dispersion and temporal characteristics of raw data (Salvucci and Goldberg, 2000). These algorithms are included in the eye tracker software and the parameters for identifying fixations can be often adjusted by the researcher.

Another essential phase of eye movements is the rapid movement of the gaze from one fixation to another, called a *saccade*. Saccades are really fast, typically lasting only 30–80 ms (Holmqvist et al., 2011). In this dissertation, saccades are loosely described as linear jumps between consecutive fixations, whereas in reality they may be of any shape. Also, we do not distinguish saccades from other kind of movements of the gaze, such as microsaccades (see e.g. Holmqvist et al., 2011), because the data set used here includes detailed information only on fixations (see Section 1.2).

In general, the main role of fixations is to gather visual information whereas the role of saccades is to move the attention and gaze to another spot. It is often thought that fixations measure allocation of attention in those locations, even though that is not always true (Holmqvist et al., 2011). However, most of the visual information is acquired during the fixations (Rayner, 2009) and therefore we will mainly focus on the fixations in the present work.

The motivation for this dissertation comes from an experiment, where the participants were looking at pictures and their eye movements were recorded. The pictures shown were static. There were no given structure which would have restricted the gaze during the experiment (as there is, for example, in reading a text), and therefore the set-up is called free-viewing condition. In general, eye movement data in this kind of free-viewing situation are rich, non-standard from the point of view of statistical analysis, and usually quite complex to analyse.

1.2 Eye movement data from an art study

The data set[1] used in this thesis was collected in an experimental study applying eye tracking for studying visual processing of paintings. Twenty students from the University of Jyväskylä

[1]The data set was collected and provided at our disposal by Pertti Saariluoma (Professor of Cognitive Science, University of Jyväskylä), Sari Kuuva (Postdoctoral Researcher, University of Jyväskylä) and exchange student María Álvarez Gil (University of Salamanca) with the technical help of Jarkko Hautala (Postdoctoral Researcher, Agora Center, University of Jyväskylä) and Tuomo Kujala (Postdoctoral Researcher, University of Jyväskylä).

participated in the experiment. Ten of the subjects were either art students or students who had studied art history and frequently visited art exhibitions, and the remaining ten subjects were students who did not have art as their major nor their hobby. The participants in the first group are here referred to as *non-novices* and participants in the latter group as *novices*.

In the experiment the participants saw six pictures of paintings and they were asked to describe orally the atmosphere in the paintings. Each painting was visible for three minutes and during the viewing their eye movements and voice were recorded. The eye movements were recorded by the SMI iView X^{TM}Hi-Speed 500 Hz eye tracker. The resolution of the screen, on which the pictures were presented, was 1024 $\times$ 768 pixels and the distance between participant's head and the screen was about 85 cm. The forehead rest was used during the experiment to minimize any extensive head movement that could reduce the quality of the recording. After the experiment the participants filled in a questionnaire about their own emotional experience on each painting.

An example of one participant's eye movements on the painting *La croisée des destins* (1988) by Risto Suomi can be seen in Figure 1.1 with the fixations during the first minute of inspection and with the 15 first saccades illustrated as arrows. The size of the painting, 766 $\times$ 768 pixels, was smaller than the screen. Eye movement data corresponding to the first ten fixations shown in Figure 1.1 are listed in Table 1.1. Each row of the table represents one fixation for which we know the time of occurrence (timestamp), duration in milliseconds and location on the screen (x- and y-coordinates in pixels).

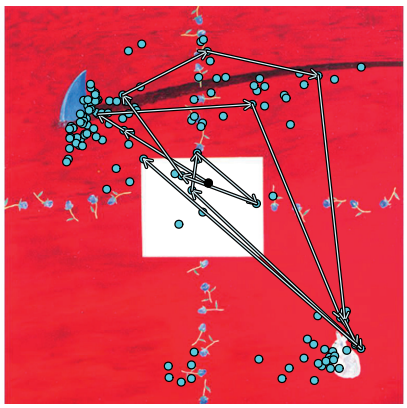

Figure 1.1: Example of the eye movements on a painting *La croisée des destins* (1988) by Risto Suomi (printed with a kind permission of the artist). The dots represent the fixations during the first minute of inspection and the arrows represent the saccades during the first five seconds. The first fixation located in the centre of the painting is marked with a black dot. The size of the painting on the screen was 766 $\times$ 768 pixels.

When extracting fixations from the raw data, the eye tracker software uses an algorithm that provides a lower bound for fixation duration. The lower bound is usually somewhere below 200 ms, for instance, in Tatler et al. (2006) the minimum fixation duration was 50 ms and in Castelhano and Henderson (2008) it was 90 ms, and it can be adjusted by the researcher. In our case, the data set included some really short fixations, but we made a decision to remove all the fixations shorter than 40 milliseconds due to the data-based assumption that they are artefacts (see Figure 4 in Article II).

Table 1.1: Example of an eye movement data matrix. Each row represents one fixation.

index	timestamp (ms)	duration (ms)	location x (px)	location y (px)
1	6	556	522	427
2	612	218	358	529
3	864	272	309	560
4	1202	198	606	579
5	1475	120	777	166
6	1631	218	816	107
7	1919	250	490	413
8	2207	396	500	485
9	2645	486	614	387
10	3181	202	470	441

1.3 Brief review of eye movement methodology in picture viewing

Since the early studies by Buswell (1935) and Yarbus (1967), researchers have tried to explain why certain elements or parts of the painting gain more fixations than some others. A fundamental contribution to this question was made by Koch and Ullman (1985), who introduced the concept of *saliency map*. A saliency map is a spatial map that is calculated from multiple feature maps of the picture and it represents the salient locations of the target picture that are assumed to attract the gaze. According to the model by Koch and Ullman (1985), the most salient parts of the picture are likely to be fixated first since, loosely speaking, they are thought to be visually attractive. The model was first implemented by Itti et al. (1998), and it has been refined after that (see further attempts to model visual attention by saliency in Borji and Itti, 2013). Saliency maps have been widely applied in the eye movement research, e.g. in Barthelmé et al. (2013); Foulsham and Underwood (2008); Parkhurst et al. (2002); Tatler et al. (2005). The weakness in most of the models relying on the saliency map is that fixation locations are assumed to be mainly driven by the features of the stimulus picture, even if there are other mechanisms influencing on where people pay attention to (see e.g. Rayner, 2009).

Apart from explaining the locations of fixations by a saliency map, there have been attempts to study the dynamics of eye movements. Noton and Stark (1971) were interested in the sample path of the gaze and introduced the term *scanpath* to describe the eye movement sequence. In their study, Noton and Stark (1971) noticed that when the subjects of an experiment were looking at the same picture of patterns over again, the first few fixations followed the same scanpath than in the first attempt. After their work, several algorithms have been presented for comparing scanpaths, such as string-edit procedures and distance algorithms, see e.g. Underwood et al. (2008) and Jarodzka et al. (2010). In general, the scanpath is a useful visualisation tool for getting a preliminary impression of the data or for data quality checking (Holmqvist et al., 2011). However, scanpath comparison algorithms are getting more and more complex by taking into account not only directions and lengths of the scanpaths but also their durations. The longer the scanpaths are, the more variation is related to them and the comparison of such scanpaths gets complicated.

Some studies argue that in the very beginning the visual stimulus is inspected very fast to

gain a general impression, and after that the subject starts to focus on local details to satisfy cognitive curiosity (see e.g. Locher, 2006; Locher et al., 1996, 2007). Furthermore, Castelhano and Henderson (2007) found in their study that the initial glimpse of the picture affects the subsequent eye movements: the picture shown quickly before the experiment had an effect on the eye movements of the actual search task. Rayner (2009) also notes that saliency is not the only factor determining where to look. Altogether, we believe that the eye movement sequence may change in time due to the visual information that is gathered during the viewing or due to some other cognitive mechanisms guiding the attention.

The analysis of eye movement data has usually been based on aggregation in time and/or in space (see e.g. Salvucci and Anderson, 1998). Aggregation summarizes the information to units which are easy to understand and compare, for instance, how many fixations lie in a certain area of interest (later AOI) or what the mean duration of fixations is in that area. In the simplest case, these AOIs can be formed by grids, but often it is the researcher who decides the size of the AOIs. The drawback of this kind of aggregation is that information on the dynamics of the process is lost. When eye movement data consist of measurements of several individuals, an experimental design usually allows groups, such as experts and novices, to be compared. Then, aggregation is done over dependent eye movement sequences and comparisons between the groups formed by independent subjects can be made using conventional statistical methods. However, there is a large variation between and within individuals regarding eye movements and hence group comparisons based on aggregated measures contain substantial uncertainty. Therefore, efficient tools for group comparisons are needed.

1.4 Point processes as models for eye movements

Statistically speaking, eye movement data can be thought to be a realisation of a spatio-temporal point process, where each point represents the location and occurrence time of an event (e.g. Diggle, 2013; Daley and Vere-Jones, 2008). This kind of data may be complex to analyse for various reasons. First, the data can be very rich when the viewing period is long since eye movements are really fast and fixations usually short. Second, eye movements gather visual information and hence they are affected by several components even in experimental studies. Such components are, for example, the target stimulus, task of the experiment or the domain knowledge of the participant (see e.g. Castelhano and Henderson, 2008; Rayner, 2009; Tatler et al., 2005). In addition, fixation pattern is usually inhomogeneous and clustered in the target picture. Therefore, the inspection of the whole eye movement sequence may be too ambitious and some aggregation may be needed. A natural choice to proceed further would be to retain the valuable information on the exact locations of fixations, but to aggregate in the temporal domain. That choice leads to spatial point process modelling and various analysis tools it offers (see e.g. Baddeley et al., 2015; Diggle, 2013; Illian et al., 2008).

Few recent studies have employed point process statistics in eye movement analysis. In their novel work, Barthelmé et al. (2013) linked the eye movements to the spatial point process framework for the first time. Their aim was to investigate how the image properties, considered using the saliency map, predict the locations of fixations. They use the inhomogeneous Poisson process as a model for fixation locations and the heterogeneous space is modelled via intensity function. In their case, the intensity is modelled as a function of the features of the image. The centrality bias is also taken into account, since people tend to look more at central locations of the pictures (partly due to the positioning of the elements in the picture). The conclusion

of the study is that the image properties do not always succeed in predicting the locations of fixations, especially when the image looks rather homogeneous. Despite the novelty and broadness of the work by Barthelmé et al. (2013), the assumption on independence of fixations is at best a rough approximation to the truth, as they note themselves.

Engbert et al. (2015) extended the idea of Barthelmé et al. (2013) and built a dynamic model for eye movements by taking the features of the picture as well as the locations of the previous fixations into account. Instead of modelling the spatial heterogeneity as in Barthelmé et al. (2013) by using the saliency map, they use the non-parametric kernel method and estimate the intensity of fixations that will serve as the *empirical saliency map* of the target picture. The idea is that the empirical saliency map describes the allocation of the attention better than a saliency map that is based on the features of the picture. For obtaining a good estimate of the empirical saliency map, Engbert et al. (2015) use the data set of all participants. Using the spatial second-order analysis they found evidence of clustering in small spatial scales that cannot be explained by the spatial heterogeneity. The dynamic model developed in Engbert et al. (2015) is able to predict this kind of spatial clustering. However, the model evaluation in Engbert et al. (2015) is based on the pair correlation function and the saccade length distribution, which do not reveal interpretable features of the temporal behaviour of the eye movement sequence. Hence there is still a lack of studies that take the temporal nature of the eye movement process into account.

According to Parkhurst et al. (2002), the saliency map can predict fixation locations only at the beginning of inspection, and then the role of the saliency diminishes. Furthermore, Barthelmé et al. (2013) show that the ability of the saliency map to predict fixation locations varied image-wise quite a lot. Engbert et al. (2015) found in their study that there is spatial clustering of fixations at short scales, which cannot be fully explained by spatial heterogeneity. All these studies then indicate that there are probably some other mechanisms that drive the attention than just the target picture represented by the saliency map or the empirical saliency map. One reasonable mechanism could be learning, which happens when the viewer gets familiar with the target picture. Also, the hypothesis of Locher (2006) about global viewing followed by a more local one supports the assumption that there might be some changes in the eye movement process during the inspection. It is rather presumable that some changes in the viewing strategy happen during the inspection. Our interest is to find out, how this kind of learning, or more generally *long-term dependence*, can be detected.

The motivation for this dissertation arose from an art study and the associated data set, but the idea of the point process approach is supported by the studies mentioned above. The work presented here contributes to the statistical inference on eye movements using spatial and spatio-temporal point process statistics. On the one hand, we expand the work of Barthelmé et al. (2013) by studying how the spatial point process statistics can be further used for fixation data. On the other hand, we follow the ideas of Engbert et al. (2015) by building dynamic models for eye movements. The first model is a simple group level model representing the average behaviour of the participants, as does the model by Engbert et al. (2015). In addition, we develop dynamic models which are sequential spatial point process models and are intended for one eye movement sequence. These subject level models are aimed for studying temporal dynamics in the sequence. The models are further extended to spatio-temporal point process framework, such that we can exploit the existing tools of spatial point processes for detecting dynamics in the sequence.

1.5 Objectives

The overall objective of this dissertation is to expand the eye movement methodology from descriptive statistics to statistical inference. This provides new summary statistics for visualisation and analysing purposes as well as modelling approaches for evaluating the uncertainty related to those summaries. The dynamic nature of eye movements is emphasised throughout the work.

The first aim of this dissertation is to find out how to approach complex eye movement data gathered in a free-viewing situation. More precisely, what kind of analysis and inference can be done without modelling and how far we can go with existing simple tools. Our solution is based on spatial point process statistics, for which a variety of methods are available. The practical motivation is to study whether participants change their inspection strategy during the viewing process. Our interest is in fixation locations and durations since they reflect which parts of the picture have been looked at and how long. We start by aggregating the spatio-temporal data over time and by studying the spatial point process of fixations, which is one of the marginals of a spatio-temporal point process. The statistical question is, how we can use such marginal processes of a spatio-temporal point process in eye movement analysis.

The second motivational question is driven by the application under study. We are interested in whether there are differences between novice and non-novice art viewers on how they are looking at paintings. This kind of group comparison is of interest in many applications using eye tracking in experimental designs. Tools are needed for the eye movement studies to improve the comparison, which is traditionally based on strong aggregation (e.g. by AOIs). Apart from the existing summary statistics for eye movements, new and justified tools are needed especially for describing how the process evolves. We introduce statistical summaries for the eye movement sequences that can be used for group comparisons as well, and study the variation related to those summaries by group level and subject level models and simulation.

The third objective is to study the evolution of an eye movement sequence and to find evidence on learning during the viewing process. This learning can appear as a change from global viewing to local one as suggested by the model in Locher (2006), or it can turn out that the gaze starts to return or avoid previously studied areas of the picture. Our interest is to find this kind of behaviour, which would statistically discriminate eye movement process from a memoryless process such as the random walk. Therefore, we develop models which remember the history of the sequence and can be used in studying the learning of the process. For simplicity, we start with a sequential spatial point process model conditioning the spatio-temporal process by the observed temporal order.

The fourth aim of this thesis is to experiment, whether we can use second-order characteristics of spatio-temporal point processes in detecting long-term dependencies such as learning in an eye movement sequence. More precisely, we are interested whether we can find statistical evidence that the spatial clustering of the fixations is confounded with the long-memory effect of fixations, thus reflecting the tendency of the gaze to return pre-examined areas of the target. For this aim, we extend the developed sequential spatial point process models to spatio-temporal models by including the occurrence times for fixations. The advantage is that then the existing spatio-temporal point process statistics and the associated software are available and can be used for statistical testing of the learning effect.

1.6 Contributions

In Article I, we study to what extent the existing methodology of spatial point process statistics can be used in the analysis of eye movement sequences and what scientific questions they can give an answer to. The issue of spatial heterogeneity has to be taken into account and here the heterogeneity is controlled by the case-control approach introduced in Diggle and Chetwynd (1991). The advantage of the case-control approach is that no detailed point process modelling is needed since we apply design-based methods for the inference. In Article I, we compare the durations of fixations within two parts of a sequence and extend the case-control approach to marked point processes to study the mark-correlation structure of two point patterns with durations as marks. The methods introduced in Article I can be used in comparing the spatial clustering structure in two point patterns, which is useful when studying whether a picture has been looked at similarly in two repeated measurements, for instance. However, by using spatial point process statistics we are not able to deduce whether this clustering is caused by the target picture or by dependencies between fixations.

The existing summary statistics of spatial point processes can be used for the preliminary analysis of the eye movement data. However, to assess the dynamics of eye movements, we need summaries that take the time or order of fixations into account. In Articles II and III we introduce summary statistics that are useful for visualising a spatio-temporal eye movement process. Due to the temporal nature of the eye movement data, some of the summaries are cumulative functions of time or order of the fixation. These summaries can be used in studying the evolution of the process, such as how fast the gaze explores the target picture, but keeping in mind their cumulative nature. Besides descriptive purposes, summary statistics are needed for testing the goodness-of-fit of the models fitted to the data.

In Article II we introduce new tools for group comparisons and a model for eye movements. Our first attempt for modelling eye movement data is to model the average behaviour of the group, in our case novice and non-novice art viewers. The model is separable such that the spatial and temporal components do not depend on each other. The developed model is used to describe the variation related to summary statistics of interest. This information is important when the statistical properties of the functional summary statistic are not known. In addition, the model can be used for group comparisons by simulating an envelope for the summary statistic under study. Furthermore, the pointwise or simultaneous envelopes of the two groups can be compared for the inference, in a similar way to confidence intervals. The model in Article II fits reasonably well to the data even though it does not change in time as the real data seem to indicate. This can be partly explained by the fact that the model is a group level model and describes the average behaviour.

Our next attempt is to understand the dynamics in a single eye movement sequence through modelling, for which we develop sequential point process models in Article III. The motivation for the modelling arises from the eye movement literature (e.g. Locher et al., 1996; Engbert et al., 2015; Tatler et al., 2006) and from our own experience that the eye movement process could contain a learning effect. The models in Article III have a so-called self-interaction component, which takes features from the whole history of the process into account. Our models apply stochastic geometry in creating self-interaction 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). The models are intended to capture such self-interaction features of an eye movement sequence. Besides the eye movement studies, these models can

be useful in other applications. In addition, sequential point process models can be a bridge to spatio-temporal models.

Finally, in Article IV we extend one of the sequential point process models of Article III to continuous time by considering the occurrence times of the fixations in addition with the locations. As a result, we get a spatio-temporal model for an eye movement sequence. The motivation for the model is to study whether the gaze has a tendency to return to the pre-examined areas of the target, the feature which we call a long-memory effect. For performing a formal test to statistically assess the existence of such an effect, we use second-order characteristics of spatio-temporal point processes. With an experiment relying on simulated data, we study in Article IV how the long-memory effect can be detected in a spatio-temporal sequence.

To sum up, the articles included in this dissertation answer to the question how spatial point processes can be used when studying the dynamics of an eye movement sequence, present a group level model that is used for assessing the uncertainty of the relevant summary statistics and also for performing group comparisons, and introduce tools for studying learning effect in an eye movement sequence by means of sequential point process models and spatio-temporal point process statistics. The model-based statistical inference relies on Monte Carlo simulation. The methods presented here are applied to an eye movement data set from an art study, but their usefulness for other applications is acknowledged.

1.7 Dissertation structure

The outline of this thesis is as follows. In Chapter 2, the underlying point process theory is introduced together with model characteristics, and the important marginals of spatio-temporal point processes are listed. Static and dynamic data summaries useful for the analysis of eye movement data in free-viewing situation are introduced in Chapter 3. Chapter 4 concerns the modelling and inference for eye movement data and Chapter 5 demonstrates developed methods in practise for the art data. Finally, Chapter 6 is for discussion.

Chapter 2

Spatial point processes and model characteristics

2.1 Spatial point process preliminaries

A *point pattern* or *point configuration* is an unordered set of points in a subset of $\mathbb{R}^d$, usually observed in a bounded observation window $W \subset \mathbb{R}^d$. For a *spatial point pattern* the dimension may be $d \geq 2$, but here we restrict it to $\mathbb{R}^2$. As an example, a set of fixations in Figure 1.1 can be considered as a spatial point pattern in the window $W = [0, 765] \times [0, 767]$.

Point processes are stochastic mechanisms that generate random point patterns. A point process X can be defined in two ways: as a random set of points, $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots\}$, or as a random *counting measure*. In the latter case, for each Borel set $B \subset \mathbb{R}^2$ the notation $N(B)$ stands for the random number of points of X that fall in the set B. N is then referred as random counting measure. For more rigorous presentations of point processes, see e.g. Chiu et al. (2013); Illian et al. (2008); Møller and Waagepetersen (2004).

A point process X is said to be

- *stationary*, if its probability law is invariant under translation: the processes X and $X + \mathbf{s}$ have the same distribution for all $\mathbf{s}$ in $\mathbb{R}^d$,
- *isotropic*, if it is invariant under rotation: the processes X and $\mathbf{r}X$ have the same distribution for every rotation $\mathbf{r}$ around the origin,
- *simple*, if no two points of X coincide, thus $\mathbf{x}_i \neq \mathbf{x}_j$ when $i \neq j$,
- *locally finite*, if for any bounded set B, the random variable $N(B)$ is almost surely finite,
- *finite*, if the total number of points in the process X is finite with probability 1.

The modelling and analysis methods developed in this dissertation are intended for points observed in a bounded window W. The observations are then modelled as a finite and simple point process. Note, however, that the underlying point process model is often assumed to be stationary.

The term *event* is used here to denote a point of the process, since each point can be attached with additional information (marks and occurrence time).

The first-order property of a point process can be described by the *intensity measure* $\mu(B) = \mathbb{E}N(B)$, $B \subset \mathbb{R}^d$. If the intensity measure μ is absolutely continuous with respect to the Lebesgue measure, then

$$\mu(B) = \int_B \lambda(\mathbf{x})\mathrm{d}\mathbf{x}, \quad B \subset \mathbb{R}^d,$$

where λ is a nonnegative function and is called the *intensity function*. The intensity function can be interpreted as

$$\lambda(\mathbf{x}) = \lim_{|d\mathbf{x}|\to 0} \frac{\mathrm{E}(N(d\mathbf{x}))}{|d\mathbf{x}|},$$

where $|d\mathbf{x}|$ is the area of a small spatial region $d\mathbf{x}$ around the location $\mathbf{x}$. For a stationary point process, $\lambda(\mathbf{x})$ is constant and interpreted as the mean number of events per unit area. The *second-order intensity function* is characterised as

$$\lambda_2(\mathbf{x}, \mathbf{y}) = \lim_{|d\mathbf{x}|,|d\mathbf{y}|\to 0} \frac{\mathrm{E}(N(d\mathbf{x})N(d\mathbf{y}))}{|d\mathbf{x}||d\mathbf{y}|}.$$

For a stationary and isotropic point process, $\lambda_2(\mathbf{x}, \mathbf{y})$ reduces to $\lambda_2(r)$, where $r = \|\mathbf{x} - \mathbf{y}\|$. Furthermore, if exists,

$$g(\mathbf{x}, \mathbf{y}) = \frac{\lambda_2(\mathbf{x}, \mathbf{y})}{\lambda(\mathbf{x})\lambda(\mathbf{y})},$$

is called the *pair correlation function*. For a stationary and isotropic point process, $g(\mathbf{x}, \mathbf{y}) = \lambda_2(r)/\lambda^2$, where $r = \|\mathbf{x} - \mathbf{y}\|$.

The point process is said to be *second-order intensity-reweighted stationary*, if the pair correlation function depends only on r and the intensity function $\lambda(\mathbf{x})$ is positive almost everywhere (Baddeley et al., 2000).

K-function

For a stationary and isotropic spatial point process one characterisation of second-order properties is provided by Ripley's K-function (Ripley, 1977), which is defined as

$$K(r) = \lambda^{-1}\mathrm{E}_o\,[\,\text{number of further events within distance } r \text{ of a typical event}\,],$$

where λ is the constant intensity and E_o is an expectation with respect to Palm probability, which heuristically means an expectation conditional on that there is a point of the process X at the origin o (considered as typical event), see e.g. Chiu et al. (2013); Illian et al. (2008). More formally, this can be written as

$$K(r) = \lambda^{-1}\mathrm{E}_o \left(\sum_{\mathbf{x}\in X\setminus\{o\}} \mathbf{1}_{b(o,r)}(\mathbf{x}) \right), \tag{2.1}$$

where $\mathbf{1}_{b(o,r)}(\mathbf{x})$ is the indicator function which equals 1 if $\mathbf{x}$ is in the ball $b(o, r)$ of radius r centred at o, and 0 otherwise. Then $\lambda K(r)$ denotes the mean number of further events within distance r of a typical event.

An important property of the K-function is that it is invariant under independent random thinning. This means that if each point is retained or discarded according to a series of mutually independent identical Bernoulli trials, then the K-function of the resulting thinned process equals to that of the original unthinned process.

The estimator of the Ripley's K-function (2.1) is recalled in Article I, see also Diggle (2013); Møller and Waagepetersen (2004). Typically, the L-function, $L(r) = \sqrt{K(r)/\pi}$, is used instead of the K-function. In the case of homogeneous Poisson process (complete spatial randomness, CSR), $K(r) = \pi r^2$ and $L(r) = r$, which makes the departures from the CSR easy to detect (Illian et al., 2008).

For a second-order intensity-reweighted stationary process, Baddeley et al. (2000) define the *inhomogeneous K-function*,

$$K_I(r) = \mathrm{E}_o \left(\sum_{\mathbf{x} \in X \setminus \{o\}} \frac{\mathbf{1}_{b(o,r)}(\mathbf{x})}{\lambda(\mathbf{x})} \right). \tag{2.2}$$

The estimator can be found in Article I and in Baddeley et al. (2000).

2.1.1 Marked spatial point processes

If a random mark, say $m_i \in M$, is attached to each event $\mathbf{x}_i \in X$, then $(X, M) = \{(\mathbf{x}_i, m_i)\}$ is called a *marked spatial point process,* where X is a point process on $\mathbb{R}^2$ and M is a mark space. In this dissertation, the mark space is limited to $\mathbb{R}^d$.

Mark-weighted K-function

For a marked spatial point process, a generalization of the K-function, called the mark-weighted K-function K_{mm}, can be defined. Following Penttinen and Stoyan (1989), we define the mark-weighted K-function as

$$K_{mm}(r) = \lambda^{-1}\mu^{-2}\mathrm{E}_o \left(\sum_{\substack{[\mathbf{x}, m(\mathbf{x})] \\ \neq (o, m(o))}} m(o)m(\mathbf{x})\mathbf{1}_{b(o,r)}(\mathbf{x}) \right), \tag{2.3}$$

where μ is the mean mark, $m(\mathbf{x}_i) = m_i$ is the mark of the event i at location $\mathbf{x}_i$ and E_o denotes the conditional expectation that there is an event with mark $m(o)$ at the origin. The estimator of K_{mm} can be found in Article I.

For separating the clustering of marks from the clustering of points, the *normalized mark-weighted K-function* is defined as

$$K_{mm}^{\mathrm{norm}}(r) = \frac{K_{mm}(r)}{K(r)}.$$

If the marks are independent, $K_{mm}(r) = K(r)$ and $K_{mm}^{\mathrm{norm}} = 1$ for all $r \geq 0$.

For a stationary and isotropic point process the mark-weighted K-function can be presented in the form (Penttinen and Stoyan, 1989)

$$K_{mm}(r) = \int_0^r \int_{\mathbb{M}\times\mathbb{M}} m_1 m_2 M_{r'}\{\mathrm{d}(m_1, m_2)\}\mathrm{d}K(r')/\mu^2. \tag{2.4}$$

Here $\mathbb{M}$ is the mark space and $M_{r'}$ is the two-point mark distribution (see e.g. Chiu et al., 2013), which is the joint distribution of two marks, $m_1 = m(\mathbf{o})$ and $m_2 = m(\mathbf{x})$, on the condition that there exists points of the process at the origin and at $\mathbf{x}$, depending only on $r' = \|\mathbf{x}\|$. When independent thinning is directed to events, the mark distribution remains the same, and hence the K_{mm}-function is also invariant under independent random thinning.

2.1.2 Multitype and multivariate point processes

If the mark space is discrete and finite, say $M = \{1, 2, \ldots, k\}$, the marked spatial point process corresponds to the *multitype (spatial) point process* $\{(\mathbf{x}_i, l_i)\}$, where $l_i \in M$ for all $i \in \mathbb{N}$. Thus the label marks l_i specify k different categories of events, such as case and control events when $k = 2$. A multitype point process is equivalent to a *multivariate point process*, which is a tuple $(X_1, X_2, \ldots, X_k)$, where each X_i is a point process consisting only of type i events (Møller and Waagepetersen, 2004).

The intensity of a subprocess X_i of a multivariate point process is $\lambda_i(\mathbf{x})$. In the stationary case, the intensity is constant λ_i. The second-order intensity of type i and type j events is

$$\lambda_{ij}(\mathbf{x}, \mathbf{y}) = \lim_{|d\mathbf{x}|,|d\mathbf{y}|\to 0} \frac{\mathrm{E}(N_i(d\mathbf{x})N_j(d\mathbf{y}))}{|d\mathbf{x}||d\mathbf{y}|}.$$

If the process is stationary and isotropic, $\lambda_{ij}(\mathbf{x}, \mathbf{y})$ depends only on the distance $r = \|\mathbf{x} - \mathbf{y}\|$.

Multitype K-function

For a stationary and isotropic multivariate point process a set of K-functions can now be defined as

$$K_{ij}(r) = \lambda_j^{-1}\mathrm{E}\left[\,\text{number of further type } j \text{ events within distance } r \text{ of a typical type } i \text{ event}\,\right],$$

where λ_j is the intensity of type j events. More formally,

$$K_{ij}(r) = \lambda_j^{-1}\mathrm{E}_{oi}\left(\sum_{\mathbf{x}\in X_j\setminus\{o\}} \mathbf{1}_{b(o,r)}(\mathbf{x})\right),$$

where E_{oi} is expectation conditional on that there is an event of type i at the origin, and X_j is the subprocess of X including only the events of type j. Then $\lambda_j K_{ij}(r)$ denotes the mean number of events of type j within distance r of a typical event of type i. It holds that $K_{ij}(r) = K_{ji}(r)$ for all $r \geq 0$, and if $i = j$ the formula gives the univariate K-function defined in (2.1). The estimator of K_{ij} can be found in Article I and e.g. in Diggle (2013).

2.2 Sequential spatial point processes

When the events $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ in a bounded window $W \subset \mathbb{R}^2$ are assigned with the order, they may be considered as a realisation of a sequential spatial point process (van Lieshout, 2006a,b). Such a sequence of events is denoted by $\overrightarrow{\mathbf{x}}_n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ here, and the corresponding sequential spatial point process by $\{(\mathbf{x}_i, i),\ i = 1, \ldots, n\}$. Examples of sequential point process constructions are e.g. random sequential absorption (RSA) models (Evans, 1993; Talbot et al.,

2000), known also as simple seq,uen,tial inhibition (SSI) in the statistical literature (Diggle et al., 1976; van Lieshout, 2006b,c). However, in their statistical use, the order is usually unobserved.

When the order of the events is known, the likelihood of a sequential point process model can be written by using conditional densities. If W^n stands for the n-dimensional space of ordered points and n is fixed, the density function f (w.r.t. the Lebesgue measure in the product space) may be defined in terms of conditional densities $f_{k+1}(\mathbf{x}|\overrightarrow{\mathbf{x}}_k)$, $k = 1, \ldots, n-1$, yielding the expression

$$f(\overrightarrow{\mathbf{x}}_n) = f_1(\mathbf{x}_1) \prod_{k=1}^{n-1} f_{k+1}(\mathbf{x}_{k+1}|\overrightarrow{\mathbf{x}}_k)\,, \tag{2.5}$$

where $f_1(\mathbf{x}_1)$ stands for the probability density of the first event and $f_{k+1}(\mathbf{x}|\overrightarrow{\mathbf{x}}_k)$ is the conditional density of a further event $\mathbf{x}$ given $\overrightarrow{\mathbf{x}}_k = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$. See Article III for more details.

2.3 Spatio-temporal point processes

For the sequential spatial point process only the time-order of the events defines the dynamics. If the waiting times between the events are also considered, we attach an occurrence time $t_i \in \mathbb{R}$ to each event $\mathbf{x}_i$. Formally, a *spatio-temporal point process* $\{(\mathbf{x}_i, t_i)\}$ is a point process in a product space $\mathbb{R}^2 \times \mathbb{R}$, and its realisation is a set of events $(\mathbf{x}_i, t_i)$ observed in a window $W \times [0, T]$. Examples of such realisation are earthquake occurrences (Ogata, 1998) or epidemics such as human *Campylobacter jejuni* infections (Gabriel and Diggle, 2009) and chickenpox incidence (Iftimi et al., 2015). These examples represent the case where typically some background process causes the births of the events. For eye movements, however, a realisation of the spatio-temporal process is similar to a particle that is moving in space and the main emphasis is on the transitions.

We consider the fixation process as a finite spatio-temporal process in $W \times [0, T]$. When stationary is used, for example in the definition of model characteristics, we think that the process is defined in $\mathbb{R}^2 \times \mathbb{R}$ and is sampled in the bounded window $W \times [0, T]$.

For a spatio-temporal point process, the first-order intensity function is interpreted as

$$\lambda_{st}(\mathbf{x}, t) = \lim_{|d\mathbf{x} \times dt| \to 0} \frac{\mathrm{E}(N(d\mathbf{x} \times dt))}{|d\mathbf{x} \times dt|},$$

where $|d\mathbf{x} \times dt|$ is the volume of an infinitesimal region $d\mathbf{x} \times dt$ containing the event $(\mathbf{x}, t)$ (see e.g. Gabriel and Diggle, 2009). If the spatio-temporal process is stationary, $\lambda_{st}(\mathbf{x}, t)$ is constant representing the mean number of events per unit area and unit time (Diggle, 2013). If the process is spatially stationary, $\lambda_{st}(\mathbf{x}, t)$ is independent of $\mathbf{x}$, and if it is temporally stationary, $\lambda_{st}(\mathbf{x}, t)$ is independent of t. The spatio-temporal process is said to be isotropic, if it is spatially isotropic and temporally reversible (Diggle, 2013).

A spatio-temporal point process is said to be *first-order separable* if the intensity can be factorised as

$$\lambda_{st}(\mathbf{x}, t) = \lambda_s(\mathbf{x})\lambda_t(t), \text{ for all } (\mathbf{x}, t) \in W \times [0, T]$$

where λ_s and λ_t are non-negative functions (see e.g. Gabriel and Diggle, 2009). Under this assumption, the effects that are non-separable can be interpreted as second-order effects rather than first-order effects. The first-order separability is assumed, for instance, when testing the hypothesis of no spatio-temporal clustering, see e.g. Diggle (2013).

The second-order intensity function is characterised as

$$\lambda_2((\mathbf{x},t),(\mathbf{x}',t')) = \lim_{|d\mathbf{x}\times dt|,|d\mathbf{x}'\times dt'|\to 0} \frac{\mathrm{E}(N(d\mathbf{x}\times dt))N(d\mathbf{x}'\times dt'))}{|d\mathbf{x}\times dt||d\mathbf{x}'\times dt'|},$$

and the pair correlation function, if exists, as

$$g((\mathbf{x},t),(\mathbf{x}',t')) = \frac{\lambda_2((\mathbf{x},t),(\mathbf{x}',t'))}{\lambda_{st}(\mathbf{x},t)\lambda_{st}(\mathbf{x}',t')}, \tag{2.6}$$

where λ_2 is the second-order intensity and λ_{st} the first-order intensity. For a stationary, isotropic process $\lambda_2((\mathbf{x},t),(\mathbf{x}',t'))$ reduces to $\lambda_2(r,u)$, where $r = \|\mathbf{x}'-\mathbf{x}\|$ and $u = |t'-t|$, and $g(r,u) = \lambda_2(r,u)/\lambda_{st}^2$.

A spatio-temporal point process is *second-order intensity-reweighted stationary* and isotropic, if the pair correlation function depends only on $r = \|\mathbf{x}'-\mathbf{x}\|$ and $u = |t'-t|$, and the intensity $\lambda_{st}(\mathbf{x},t)$ is bounded away from zero (Gabriel and Diggle, 2009).

Spatio-temporal K-function

For a stationary and isotropic process, the spatio-temporal K-function is defined as

$$K_{st}(r,u) = \lambda_{st}^{-1}\mathrm{E}_o\,[\,\text{number of further events within distance } r \text{ and time } u \text{ of a typical event}\,],$$

where E_o is the expectation on condition that there is an event of the process at the origin at time 0. More formally,

$$K_{st}(r,u) = \lambda_{st}^{-1}\mathrm{E}_o\left(\sum_{\substack{[\mathbf{x},t(\mathbf{x})]\\ \neq(o,0)}} \mathbf{1}_{b(o,r)}(\mathbf{x})\mathbf{1}_{[-u,u]}(t(\mathbf{x}))\right), \tag{2.7}$$

where $t(\mathbf{x}_i) = t_i$ is the time of the event i at location $\mathbf{x}_i$. Note, that Gabriel and Diggle (2009) consider only future events, whereas we take into account both past and future events.

The spatio-temporal K-function can be presented in the form

$$K_{st}(r,u) = \int_0^r \int_{\mathbb{T}\times\mathbb{T}} \mathbf{1}_{[-u,u]}(t_2-t_1)F_v\{\mathrm{d}(t_1,t_2)\}\mathrm{d}K(v),$$

where $\mathbb{T} = [0,T]$ and F_v is the joint cdf of the occurrence times of two events, one at the origin and the other one at $\mathbf{x}$ with $\|\mathbf{x}\| = v$. Here $\mathbf{1}_{[-u,u]}(t_2-t_1) = 0$ if the time between occurrences of two events longer than u, 1 otherwise. In analogy with the case of mark-weighted K-function in (2.4), one can deduce that spatio-temporal K-function is invariant under independent random thinning.

For a second-order intensity-reweighted stationary and isotropic point process, Gabriel and Diggle (2009) define the *space-time (or spatio-temporal) inhomogeneous K-function* (STIK-function) as

$$K_{st,I}(r,u) = 2\pi \int_{-u}^{u} \int_{0}^{r} g(r',u') r' dr' du', \tag{2.8}$$

where g is the pair correlation function in (2.6). See, for instance, Article IV or Gabriel and Diggle (2009) for the estimator.

Mark-weighted spatio-temporal K-function

We may define a mark-weighted spatio-temporal K-function for a stationary and isotropic process by

$$K_{mm,st}(r,u) = \lambda_{st}^{-1} \mu^{-2} \mathrm{E}_o \left(\sum_{\substack{[\mathbf{x},m(\mathbf{x}),t(\mathbf{x})] \\ \neq (o,m(o),0)}} m(o) m(\mathbf{x}) \mathbf{1}_{b(o,r)}(\mathbf{x}) \mathbf{1}_{[-u,u]}(t(\mathbf{x})) \right), \tag{2.9}$$

where E_o is the expectation conditioning that there is an event of the process with mark $m(o)$ at the origin at time 0. If the marks are uncorrelated, the function reduces to (2.7), i.e. $K_{mm,st}(r,u) = K_{st}(r,u)$.

The mark-weighted spatio-temporal K-function can now be presented as

$$K_{mm,st}(r,u) = \int_0^r \int_{\mathbb{M}\times\mathbb{M}} \int_{\mathbb{T}\times\mathbb{T}} \mathbf{1}_{[-u,u]}(t_2 - t_1) f(m_1,m_2) F_{r'}\{\mathrm{d}(t_1,t_2)\} M_{r'}\{\mathrm{d}(m_1,m_2)\} \mathrm{d}K(r'),$$

where $\mathbb{T} = [0,T]$, $F_{r'}$ indicates the joint cdf of the birth times of two events, one at the origin and the other one at $\mathbf{x}$ with $\|\mathbf{x}\| = r'$, and $M_{r'}$ is the corresponding two-point mark distribution. This is also invariant under independent random thinning.

Conditioning on the past

The characterisation of the spatio-temporal point process as a point process in the product space $W \times [0,T]$ allows, for example, the study of spatio-temporal clustering. However, the moment characteristics of such a process do not capture dependence on the past. In addition, the likelihood is seldom possible to present in a closed form. These problems can be partially bypassed by utilising the temporal ordering of points which leads to the conditional densities with respect to the history of the process. We assume here that all the finite-dimensional densities exists, that is, they are absolute continuous with respect to the Lebesgue measure of the same dimension (the point process is regular in the terminology of Daley and Vere-Jones, 2003).

Alternative way to characterise a spatio-temporal point process is to consider a point process in time and treat the locations as marks. This kind of process has a causal feature and conditioning by the past is natural. In that case, a spatio-temporal point process is typically defined in terms of conditional intensities, since the construction aims to explain how the evolution of the process depends on its history.

The conditional intensity function of a spatio-temporal point process is the spatial first-order intensity at time t conditional on the history $\mathcal{H}_{t-}$. The history of the process refers to the events that occur before time t, thus $\mathcal{H}_{t-} = \{(\mathbf{x}_i, t_i),\, t_i < t\}$ (Diggle, 2013). Here we assume that the process is *orderly*, which implies that $t_i < t_{i+1}$ for all $i \geq 1$ and no multiple events occur. The conditional intensity of a spatio-temporal point process can be defined informally as

$$\lambda^*(\mathbf{x}, t|\mathcal{H}_{t-}) = \lim_{|d\mathbf{x}\times dt|\to 0} \frac{\mathrm{E}(N(d\mathbf{x}\times dt)|\mathcal{H}_{t-})}{|d\mathbf{x}\times dt|}.$$

For an orderly spatio-temporal point process, the conditional intensities define the process uniquely (Diggle, 2013). As González et al. (2016) note, the likelihood-based inference is quite straightforward for any model for which the conditional intensity is specified. For a spatio-temporal point process on a finite region $W \times [0, T]$ this requires that $\lambda^*(\mathbf{x}, t|\mathcal{H}_{t-})$ is non-negative and finite for all $(\mathbf{x}, t) \in W \times [0, T]$ and for all possible configurations $\mathcal{H}_{t-}$ with $t \in [0, T]$ (Diggle, 2013).

Likelihood inference

Let us assume that the point process $\{t_i\}$ with marks $\{\mathbf{x}_i\}$ is regular (see e.g. Daley and Vere-Jones, 2003). For a set of events, $(\mathbf{x}_i, t_i)$, $i = 1, \ldots, n$, observed in $W \times [0, T]$, the log-likelihood function can be written as (see e.g. Diggle, 2013)

$$\sum_{i=1}^{n} \log \lambda^*(\mathbf{x}_i, t_i|\mathcal{H}_{t_i-}) - \int_0^T \int_W \lambda^*(\mathbf{x}', t'|\mathcal{H}_{t-})\mathrm{d}\mathbf{x}'\mathrm{d}t',$$

where $\mathcal{H}_{t_i-}$ denotes the history up to time t_i. Note that the conditional intensity may be intractable, as is the case for the log-Gaussian Cox process, and calculation of the integral may be complicated (Diggle, 2013; González et al., 2016). An alternative for this approach is a partial likelihood familiar in the context of proportional hazard modelling of survival data, which is applied to spatio-temporal setting in Diggle et al. (2010). However, as González et al. (2016) argue, it is sometimes more suitable to treat spatio-temporal point processes as a random collection of events, as they are defined, observed in some region where one of the dimensions represents time. This approach emphasises the spatial nature of the process but employs the natural ordering of time (González et al., 2016).

Our interest is in the spatial locations of the events whilst the temporal process is considered to be less important. The temporal part of the eye movement sequence gives a natural ordering of the points and allows us to write the likelihood for the process in a sequential way. By conditioning on the number of events n, we have constructed a sequential point process model for the locations in (2.5). Under the regularity assumption (see Daley and Vere-Jones, 2003) and assuming that the spatial transitions do not depend on the previous occurrence times, the model can be extended to the spatio-temporal point process framework (see Section 4.2.1 and Article IV). In this case, we consider locations $\mathbf{x}_i$ as exogenous variables, which affect the process of occurrence times (see e.g. Daley and Vere-Jones, 2003). Therefore, for modelling the locations $\overrightarrow{\mathbf{x}}_n$, we can use the likelihood of the sequential point process model, whereas for modelling the occurrence times, we need the conditional distributions of t given the history that includes both occurrence times and locations of the events in the past.

2.4 Useful marginals of spatio-temporal point processes

Eye movement data have been analysed in terms of spatial point processes (see Section 1.4), even though the phenomenon can be considered as spatio-temporal. This is justified in some cases, as Diggle (2013, p. 195) writes: "*Most spatial processes in nature are merely snapshots of evolving spatio-temporal processes. But to argue that they should therefore be analysed using spatio-temporal methods would be misguided. We analyse purely spatial data, and build purely spatial models, if and when in doing so we can address interesting scientific questions.*"

We apply spatial point processes to get information on the underlying spatio-temporal point process. The question is, if the other marginals of spatio-temporal point processes could be employed as well. Next, we will introduce some marginals of spatio-temporal point processes that can be considered relevant for eye movement analysis.

Assume we have a finite marked spatio-temporal point process, where each event includes information on the location and occurrence time and is attached with a mark vector that consists of the duration and label marks. More precisely, consider a spatio-temporal point process, where the events result from one particle (e.g. the gaze) moving in space such that no simultaneous events occur. The particle arrives in location $\mathbf{x}_i$ at time t_i, stays there for time m_i and arrives the next location $\mathbf{x}_{i+1}$ at time t_{i+1}. A duration mark m_i is needed if the movement of the particle from one location to another takes considerable amount of time and the duration cannot then be approximated by $t_{i+1} - t_i$. Relevant marginal processes which can be derived from such marked spatio-temporal point process $\{(\mathbf{x}_i, t_i, \mathbf{m}_i), i = 1, \ldots, n\}$ are the following ones:

- The (unordered) spatial point process $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in W, which indicates where the events are located.
- The sequential spatial point process $\{(\mathbf{x}_i, i), i = 1, \ldots, n\}$ in $W \times \{1, \ldots, n\}$ takes the order of events into account and reveals the trace of the process.
- The (ordered) temporal point process $(t_1, t_2, \ldots, t_n)$ in $[0, T]$ reveals the occurrence times of the events, and can be used to study if the temporal behaviour of the process changes in time.
- The (unordered) marked spatial point process $\{(\mathbf{x}_i, m_i), i = 1, \ldots, n\}$ in $W \times M$, where m is the duration and $M \subset \mathbb{R}$ is the mark space, associates each event with its duration. This can be used when studying in which locations the particle is staying longer than on average.
- The ordered marked spatial point process $\{(\mathbf{x}_i, m_i, i), i = 1, \ldots, n\}$ in $W \times M \times \{1, \ldots, n\}$ can be used to study how the process evolves in space and whether the marks change in time, for example, are there long jumps and short events in the beginning of the process.
- The multitype spatial point process $\{(\mathbf{x}_i, l_i), i = 1, \ldots, n\}$ in $W \times M$, where l is the label mark and $M = \{1, \ldots, k\}$. The label can indicate
  - subinterval U_j of the partition of $[0, T]$ to which the event belongs. Then we can compare the subprocesses of events in two time intervals.

– subspace W_j of the partition of W to which the event belongs. These W_j:s can be considered as areas of interests (AOIs) and this approach can be used to study subprocess of events in an AOI.

In an eye movement framework, the eye movement process can be considered as a marked spatio-temporal point process, of which events represent fixations. Each fixation has a location $\mathbf{x}_i$, time of occurrence t_i and duration m_i as a mark. When the order of fixations is known, it is possible to calculate the lengths of the jumps (i.e. saccades, if they are considered as line segments), $||\mathbf{x}_{i+1} - \mathbf{x}_i||$, as well as the empirical jump length distribution. In addition, the trace of the gaze, called scanpath (see Chapter 3), can be calculated for sequential spatial or spatio-temporal sequences.

The choice of an appropriate marginal process is determined by the research hypothesis in question. In eye movement studies, Barthelmé et al. (2013) use spatial point processes, whereas the model in Engbert et al. (2015) is a spatio-temporal point process model, but their focus is in spatial characteristics and interactions, whereas the temporal nature is ignored. In this dissertation, we concentrate on spatial, sequential spatial and spatio-temporal point processes, but also unordered marked spatial and multivariate spatial point processes are applied. Altogether, the locations of fixations are always considered and usually the temporal information either as order or as occurrence time of the fixation is also taken into account.

Chapter 3

Data summaries

The information in data is typically compressed into a summary statistic which describes the feature of interest. However, any data summary alone is not very informative without knowledge about the statistical variation in it. In this dissertation, we are interested in describing features of the eye movement data with existing and new summary statistics. The statistical variation in the summary can be acquired by modelling and parametric bootstrap method (Efron and Tibshirani, 1994, p. 53): when simulating the fitted model we get an impression of the sampling distribution of the summary statistic in question and we can then evaluate its variation. In addition to descriptive purposes, summary statistics are useful in validating the fitted models and in group comparisons (see e.g. Article II).

In this chapter, we first introduce some commonly used static summaries for eye movement data. Then we concentrate on functional summary statistics that describe the dynamic nature of the eye movement process. Both kind of summaries are useful in visualisation and modelling. Additional summary statistics and several other eye movement measures are listed in Holmqvist et al. (2011) with references to literature.

Here, we introduce the notation for the following chapters considering eye movements as spatio-temporal point processes. Assume that a participant of an eye movement experiment is inspecting a target picture on a screen for fixed time, say T milliseconds. The eye tracker records movements of the participant's gaze in a given bounded window $W \subset \mathbb{R}^2$, representing the target picture, during the inspection time interval $[0, T]$. The obtained fixation sequence can be interpreted as a set $\{(\mathbf{x}_i, t_i, m_i),\ i = 1, \ldots, n\}$, where $\mathbf{x}_i \in W$ is the location, $t_i \in [0, T]$ is the occurrence time and $m_i \in \mathbb{R}^+$ is the duration of the fixation i. This kind of sequence can be considered to be a realisation of a simple (marked) spatio-temporal point process.

3.1 Static data summaries

Recall that the locations where the gaze stops have often been of interest in eye movement studies (see Section 1.3). The simplest data summary that describes the spatial distribution of fixations is the *point map*, a graphical presentation of the point pattern of fixations. For exploring an eye movement data, plotting a point map is the first thing to do. It represents which parts of the picture have been looked at and how the fixations are located with respect to each other. However, especially when doing comparisons, it is more useful to deal with functions rather than point patterns, and hence *regionalization* or *smoothing* of a point pattern

comes into use. Both of these methods employ the point pattern data and result in a two-dimensional function (Illian et al., 2008). For example, by using the kernel smoothing method, we can estimate the underlying spatial intensity $\lambda(\mathbf{x})$ based on a point pattern. As a result, we get an estimate of the intensity for each point $\mathbf{x} \in W$, which allows, for instance, statistical comparison of two intensity surfaces (see Article II). The eye tracker software usually provides a heatmap visualisation, which resembles smoothed point map but includes weighting by fixation durations.

The duration of fixation is a commonly used measure for eye movement data. In eye movement studies, often the mean and standard deviation of fixation duration are reported, even thought the empirical distribution is usually right-skewed as is common for many lifetime measures. Here, we suggest to study the distribution rather than aggregated summaries of fixation duration and, in case of group comparisons, to compare the distributions by means of the shift function (Doksum and Sievers, 1976) measuring how much one distribution has to be shifted in order to coincide with the other. This comparison reveals if one of the groups has longer or shorter fixations more than the other group, and is more informative than the difference of the empirical means. Examples of the use of the shift function can be found in Article II.

Another distribution that is of interest is the jump length distribution. Here, the jump refers to a saccade, defined as a line segment between two consecutive fixations. The length of the jump between the subsequent fixations i and $i+1$ is then $||\mathbf{x}_{i+1} - \mathbf{x}_i||$. In the eye movement literature, the jump length is usually called *saccade amplitude* and measured in degrees of visual angle (see e.g. Holmqvist et al., 2011). Typically, the gaze makes short jumps while the subject is concentrating on details or after a long jump when adjusting the gaze to the target of interest. When the stimulus target of an eye movement study is represented on a computer screen, the jumps are restricted such that no jumps outside the observation window are allowed. In practice, this means that the fixations outside the window are excluded, as well as the saccades leaving or entering the observation window. Thus, there is always a maximum distance for a jump length: the diameter of the observation window. The jump length is a useful measure, for example, when comparing eye movements in two different tasks (see e.g. Tatler et al., 2006).

3.2 Dynamic data summaries

For describing how the eye movement process evolves, dynamic data summaries measured as a function of time should be employed. To begin with, the point maps can be presented as dynamic summaries of spatial dispersion considering time. A *dynamic point map* can be drawn such that the colour of the point indicates its order (see e.g. Figure 2 in Article III). Alternatively, the order of the points could be denoted by connecting them with arrows, as in Figure 1.1. This dynamic visualisation represents the scanpath of the gaze.

Goldberg and Kotval (1999) defined the *scanpath length* as the total distance that the gaze travels during the inspection period. Usually, the scanpath is thought to be a sequence of consecutive fixations connected with the line segments (saccades) between them, and hence the scanpath length can be calculated as a sum of jump lengths. Scanpaths can be used for visualising the movement of the gaze (see Figure 1.1), but they have also been used for comparisons of subjects or viewing processes e.g. by string-edit methods mentioned in Section 1.3. The string-edit methods require that the target window is divided into the areas of

interest (AOIs). If the scanpath is very long, the string that consists of the AOI labels can also be long and comparing two such strings is perhaps not meaningful. For long sequences, another option would be to define AOIs and estimate the *transition probability matrix* that consists of the probabilities for the jumps of the process between the AOIs. The transition probability matrix can then be used to reveal the most commonly occurring transitions. Transition probabilities are used in testing the goodness-of-fit of the models in Article II.

We can apply the functional version of the scanpath length measure defined by Goldberg and Kotval (1999) in order to study the movements of the gaze. The scanpath length is then calculated for each fixation occurrence time, whether it is a discrete (sequential spatial point process) or continuous (spatio-temporal point process). Let us denote the ordered set of fixations up to a fixed time t as $x = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_t}\}$. The scanpath length for this point set is

$$S_P(t) = \sum_{i=1}^{n_t - 1} ||\mathbf{x}_{i+1} - \mathbf{x}_i||, \tag{3.1}$$

where $\mathbf{x}_i$ is the location of the fixation i. The resulting scanpath length curve increases rather fast if the gaze takes long jumps, for example, during the global view in the beginning of the inspection.

Another useful summary statistic, which was first introduced by Goldberg and Kotval (1999), is based on the convex hull. Let us assume that the observation window W is convex. The *convex hull* of the point set $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_t}\}$ is defined as the minimal convex set of W which contains all the (unordered) points $(\mathbf{x}_1, \ldots, \mathbf{x}_{n_t})$. More formally, convex hull is the set of all convex combinations of $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_t}\}$,

$$C_x(t) = \left\{ \sum_{i=1}^{n_t} \alpha_i \mathbf{x}_i : \alpha_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{n_t} \alpha_i = 1 \right\}. \tag{3.2}$$

An example of the convex hull is shown in Figure 3.1, left. We measure the convex hull as a function of time t and define the *convex hull coverage* as the relative area of the convex hull,

$$S_C(t) = \begin{cases} 0 & \text{if } n_t < 3 \\ |C_x(t)|/|W| & \text{if } n_t \geq 3. \end{cases} \tag{3.3}$$

The convex hull coverage is then a cumulative function and it is most useful for measuring the coverage of the target window at the beginning of the process. The problem with this measure is that if the gaze makes long jumps towards the edges of the spatial window W, then the convex hull increases rapidly and do not reveal detailed information on the process later on since the following fixations will most probably appear inside it.

Let us define the ball union measure of points $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ as the union of the balls $b(\mathbf{x}, r)$ with radius r located at each point $\mathbf{x}$ (see Figure 3.1, left). The relative area of this measure inside the window W is called the *ball union coverage*, defined as

$$S_B(t) = \frac{|\bigcup_{i=1}^{n_t} b(\mathbf{x}_i, r) \cap W|}{|W|}. \tag{3.4}$$

It is a monotonically increasing function and each new fixation increases the ball union coverage by πr^2 at the maximum. The ball union coverage depends on the radius r, which must be chosen by the researcher. The radius can be adjusted, for example, with respect to the size of the foveal region, in which the visual image is accurate.

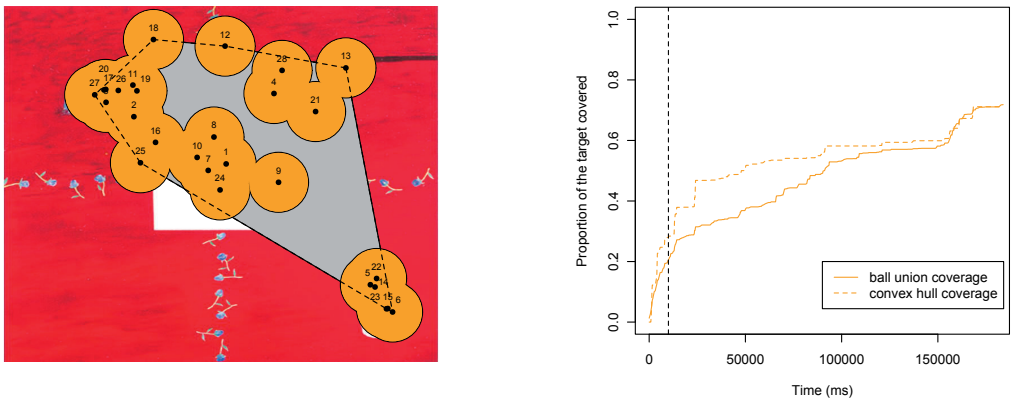

Figure 3.1: Left: Example of the convex hull and ball union coverage measures of fixations during the first 10 seconds of inspection. The points represent locations of the fixations and the numbers indicate their order. The grey polygonal area is the convex hull coverage and the orange area is the ball union coverage (with radius $r = 50$ pixels). Right: Convex hull and ball union coverage functions for the fixations during three minute inspection period. The dashed vertical line indicates the 10 second time point.

Both coverage measures, convex hull and ball union, are calculated with respect to the size of the window W, and hence they receive values between $[0, 1]$ (or 0–100%), see Figure 3.1, right. The advantage of these coverage measures is that they do not require that the space is divided into the areas of interest or into a grid.

Recall that we are interested in learning in an eye movement sequence. More precisely, how the already existing fixations attract or inhibit forthcoming fixations. For this we suggest *recurrence measure*, which calculates the number of earlier fixations $\overrightarrow{\mathbf{x}}_k = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$ closer than radius r of the current fixation $\mathbf{x}$, formally, $\sum_{i=1}^{k} \mathbf{1}_{b(\mathbf{x}_i,r)}(\mathbf{x})$. The delayed version of this measure omits the previous fixation $\mathbf{x}_k$ and is further preferred in this dissertation. We define a *cumulative delayed recurrence measure* for a point set $x = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_t}\}$ as

$$S_R(t) = \begin{cases} 0 & \text{if } n_t < 3 \\ \sum_{i=3}^{n_t} \sum_{j=1}^{i-2} \mathbf{1}_{b(\mathbf{x}_j,r)}(\mathbf{x}_i) & \text{if } n_t \geq 3. \end{cases} \tag{3.5}$$

This measure indicates how much the process favours locations nearby the previous fixations. The recurrence measure is employed in the sequential point process model, introduced in Article III, such that the process favours locations close to the previous fixations. Illustrations of the cumulative delayed recurrence measure can be found in Article III and in Chapter 5.

The data summaries presented above are found to be useful for eye movement data both in the literature and in the articles included in this dissertation. In addition to those summaries, empirical versions of the point process characteristics can be employed. The second-order characteristics presented in Chapter 2 can be used for studying the interactions between the events. When describing the spatial clustering of fixations pair correlation functions come into use, whereas for testing the clustering K-functions are usually applied. For these second-order measures, there are also inhomogeneous, marked and spatio-temporal versions available, see Chapter 2.

Chapter 4

Modelling and inference for eye movement data

Eye movement sequences may be difficult to model in terms of spatio-temporal point process models due to non-stationarity of long sequences or lack of information in short sequences. Therefore, we first approach the analysis of an eye movement data by design-based methods, which are based on marginals of a spatio-temporal point process and do not require any parametric modelling. Design-based in this context means that we condition on the locations of fixations and the statistical test against the null hypothesis can be formulated by random labelling. Random labelling is based on permutation of the labels of the points or the subjects (see more in Articles I and II). Design-based methods that are presented here are easy to apply using existing point process statistics, such as the K-function. However, the usefulness of design-based methods for ordered point processes is limited and they can answer to rather simple research questions. Up to our knowledge, modelling is required for assessing the uncertainty in functional summary statistics or for studying long-term dependence in an eye movement sequence.

4.1 Model-free approaches

One of the fundamental questions in the eye movement research is, to which parts of the target picture the subjects pay attention. If we assume that fixations exhibit the allocation of attention, this can be easily visualised by plotting the point map of fixations. However, if the question is whether the eye movement patterns of two groups are similar, some smoothing of the point map is needed and, for example, the intensity function can be applied. The intensity surface (the estimate of the intensity function $\lambda(\mathbf{x})$ in two-dimension) can be used to visually summarize, how the fixations are spread on a target picture. In group comparisons, it is interesting to compare two intensity surfaces statistically in order to find out whether the two groups pay attention to the same parts of the target picture. This method is described in Article II, where the intensity surfaces of novices and non-novices are compared. The test statistic can be based on the ratio of the estimated intensities for the two groups. The statistical comparison can be done by random labelling of subjects into two groups several times and by calculating the ratio of the intensity surfaces for the new groups in each permutation. The random labelling of subjects retains the fixation sequence as full, which means that the

individual effect, if there is any, is preserved.

The K-function is a second-order characteristic of a spatial point process which measures the clustering of events as a function of distance. However, the K-function assumes stationarity, which is a rather strong assumption in the case of eye movement data. To deal with the spatial heterogeneity, one possibility is to apply the case-control design introduced by Cuzick and Edwards (1990) and used in epidemiological studies. The approach is based on the idea that we choose a group of controls, which is thought to be a random sample from the underlying population. Then, the point pattern of cases (e.g. the ones with the disease) can be tested with respect to the point pattern of controls in order to study *extra (spatial) clustering* of cases. More precisely, we can test whether cases are more clustered than controls by comparing the difference of the two K-functions to an envelope created by random labelling of the bivariate points. An exact Monte Carlo test can be done by using a test statistic, see e.g. Diggle (2013); Diggle and Chetwynd (1991). As an example, in Article I we study the dynamics of the fixation process in a rough way by comparing how fixations are clustered in the beginning versus in the latter part of the inspection period.

The two design-based methods discussed above are intended for spatial point patterns. However, the clustering of fixations on a picture is not only a cause of its features, but probably also of other mechanisms related to visual processing as noted earlier. By using spatial point process statistics, we do not get detailed information on the temporal evolution of the process, such as whether the gaze has returned to some elements of the picture multiple times during the process, or whether the clustering is only due to consecutive fixations. To answer these questions, we cannot base the inference on the spatial point process marginal of the spatio-temporal point process, but at least the temporal order of fixations is needed. The problem is that the similar case-control method that was used for spatial point patterns is not available for sequential spatial point patterns, because events in the ordered sequence cannot thought to be exchangeable.

The design-based method can be used to test the lack of spatio-temporal interaction in a spatio-temporal sequence, that is, whether the spatio-temporal K-function decomposes into spatial and temporal components, such that $K(\mathbf{s}, t) = K(\mathbf{s})K(t)$ is true. The test is based on random labelling the locations, but keeping the occurrence times fixed (see e.g. Diggle et al., 1995). Here the random labelling suffices, since the idea is to find out whether the temporal and spatial domain interact or not.

Another commonly tested hypothesis for a spatio-temporal point process is spatio-temporal clustering. This test, however, relies on modelling through the null hypothesis that the underlying process is a Poisson process. The test is performed by comparing the empirical spatio-temporal K-function to K-functions calculated for simulations from an inhomogeneous Poisson process with intensity $\hat{\lambda}_{st}(s, t) = \hat{\lambda}(s)\hat{\lambda}(t)$ (Gabriel and Diggle, 2009). The test statistic is then based on the difference between the empirical K-function and its expectation under the null hypothesis (see e.g. Diggle, 2013). Nevertheless, this test or any of the model-free approaches presented above are not able to detect long-term dependency that causes spatial clustering in a long time range in an eye movement sequence. Article IV addresses to this issue by modelling and applying the spatio-temporal inhomogeneous K-function.

To conclude, modelling is needed for sequential spatial or spatio-temporal sequences since the available design-based methods do not give answers to the research questions we have set.

4.2 Dynamic models

Our interest is to study what happens inside the eye movement sequence and hence in modelling we concentrate on sequential spatial and spatio-temporal point processes. Important characteristic of eye movements, such as spatial inhomogeneity and dependency between fixations, should be taken into account in the modelling. The model presented by Barthelmé et al. (2013) does not fulfil these requirements, because it assumes independence of fixations and it is a spatial point process model. The dynamic model by Engbert et al. (2015) considers the spatial heterogeneity and second-order spatial interaction, but the model is rather complex and is validated only with static summaries that do not take the time into account. We propose new point process models that can be used for any eye movement data collected during free-viewing and where the target picture is static and do not have strict demands for the movement of the gaze (such as linear movement while reading).

4.2.1 Spatio-temporal point process model for fixations

The density of a regular spatio-temporal point process $\{(\mathbf{x}_i, t_i)\}$ can be defined by two families of conditional probability densities,

$$\{f_k(\mathbf{x}|\overrightarrow{t}_{k-1}, \overrightarrow{\mathbf{x}}_{k-1})\} \text{ and } \{h_k(t|\overrightarrow{t}_{k-1}, \overrightarrow{\mathbf{x}}_{k-1}, \mathbf{x_k})\},$$

where $k \in \mathbb{N}$ and $\overrightarrow{\mathbf{x}}_k$ and $\overrightarrow{t}_k$ denote the ordered sequences of fixation locations and fixation occurrence times, respectively. The density $f_k(\mathbf{x}|\overrightarrow{t}_{k-1}, \overrightarrow{\mathbf{x}}_{k-1})$ describes the distribution of the location of the kth event given the history up to time t_{k-1} and the density $h_k(t|\overrightarrow{t}_{k-1}, \overrightarrow{\mathbf{x}}_{k-1}, \mathbf{x}_k)$ describes the distribution of the kth occurrence time given the history of the process up to time t_{k-1} and the location of the kth point. Similar construction of a spatio-temporal point process is presented in Jensen et al. (2007), but we emphasise here the locations instead of the occurrence times.

From now on, let us assume that the data consist of pairs $(\mathbf{x}_i, t_i)$, $i = 1, \ldots, n$, where $t_i \in [0, T]$ and $\mathbf{x}_i \in W \subset \mathbb{R}^2$ for all $i = 1, \ldots, n$. The joint density of $(\overrightarrow{\mathbf{x}}_n, \overrightarrow{t}_n)$, can be presented in general form as

$$f(\overrightarrow{\mathbf{x}}_n, \overrightarrow{t}_n) = f_1(\mathbf{x}_1, t_1) \prod_{k=2}^{n} f_k(\mathbf{x}_k, t_k|\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1}),$$

where f_1 is the model for the occurrence time and location of the first fixation, or is omitted if $(\mathbf{x}_1, t_1)$ is assumed to be fixed, and $f_k(\mathbf{x}_k, t_k|\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1})$ is the density for a new point $(\mathbf{x}_k, t_k)$ conditional on the history up to time t_{k-1}. We can decompose the density as

$$f_k(\mathbf{x}_k, t_k|\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1}) = f_k(\mathbf{x}_k|\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1}) h_k(t_k|\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1}, \mathbf{x}_k),$$

see also Jensen et al. (2007).

We think that the order of fixations include enough information on the temporal dynamics of the eye movement process for our purposes and the occurrence times of fixations would bring only minor value for the analysis. Therefore, we put our emphasis first on the modelling of the ordered sequence of fixation locations before considering their occurrence times. This approach challenges the traditional way of treating spatio-temporal point processes as temporal point processes with locations as marks and modelling them with conditional intensities.

Next, we make two simplifying assumptions, which support the idea that we condition by the locations of fixations and model them using sequential spatial point process models. First, we assume that the number of points, n, is fixed. This is reasonable if the number of fixations is not very informative for long eye movement sequences, as we believe. Second, we assume that the spatial transitions $\mathbf{x}_{k-1} \to \mathbf{x}_k$ given the previous locations $\overrightarrow{\mathbf{x}}_{k-1}$ do not depend on the occurrence times $\overrightarrow{t}_{k-1}$, i.e. $f_k(\mathbf{x}_k|\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1}) = f_k(\mathbf{x}_k|\overrightarrow{\mathbf{x}}_{k-1})$ for all $k = 2, \ldots, n$. Due to this assumption, we can fit a sequential point process model for the sequence of fixation locations $\overrightarrow{\mathbf{x}}_n$, and then a temporal model for the occurrence times $\overrightarrow{t}_n$.

Under these assumptions, we can write the joint density as

$$f(\overrightarrow{\mathbf{x}}_n, \overrightarrow{t}_n) = f(\overrightarrow{\mathbf{x}}_n)h_1(t_1) \prod_{k=2}^{n} h_k(t_k|\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1}),$$

where $f(\overrightarrow{\mathbf{x}}_n) = f(\mathbf{x}_1)\prod_{k=2}^{n} f_k(\mathbf{x}_k|\overrightarrow{\mathbf{x}}_{k-1})$ is a sequential point process model as in (2.5), and h_k models the occurrence times $\overrightarrow{t}_k$ conditional on $(\overrightarrow{\mathbf{x}}_{k-1}, \overrightarrow{t}_{k-1})$. In this construction, the role of fixation locations $\overrightarrow{\mathbf{x}}_{k-1}$ when modelling the occurrence times t_k is to be treated as covariates (Daley and Vere-Jones, 2003). A simplification of this kind of spatio-temporal point process model is demonstrated in Article IV.

4.2.2 Sequential spatial point process models for fixations

This subsection introduces sequential spatial point process models indented for extracting features from a fixation sequence measured from one subject (see also Article III). The models consist of three components, which are thought to catch important characteristics of eye movement sequences. The reasoning of the components and their construction for the models are reviewed next. The models can be extended to spatio-temporal point process models by adding occurrence times of fixations as explained above.

Spatial heterogeneity

The first feature of an eye movement sequence that the model should take into account is *spatial heterogeneity* caused by the target picture. Spatial heterogeneity appear in any eye movement data, because people do not look at pictures uniformly. There have been attempts to explain locations of fixations by saliency maps (see e.g. Barthelmé et al., 2013, and Section 1.3). However, saliency maps, being models of visual attention, are intended to predict fixation locations in the beginning of inspection and they do not always perform well (see Barthelmé et al., 2013). Another estimate for the spatial heterogeneity in the case of eye movements is suggested by Engbert et al. (2015). In their study, the spatial heterogeneity is described by empirical intensity function (also known as empirical saliency map) that is estimated using pooled fixation data from all the subjects participating the experiment. The empirical intensity function, however, does not measure exactly the same than the saliency map: when the saliency map is calculated from the features of the picture and is representing the areas which should receive attention, the intensity surface describes which areas of the picture have been looked at in general and how intensively.

In our data set, we have 20 participants and hence we have 20 independent replicates of the eye movement sequence per picture, which can be used for estimating the empirical

intensity. The reason for employing replicates is that the intensity surface describes the first-order characteristic of the point pattern and in order to study second-order or higher-order characteristics we have to use some additional data (e.g. covariates or replicates as suggested by Diggle et al., 2007). The drawback of this empirical intensity approach is that the short saccades are more common than the long ones and hence there might be extra clustering which the features of the picture alone cannot explain, as noted by Engbert et al. (2015). To reduce the effect of dependency between the fixations one could use thinned sequences, for example to choose only every 4th fixation of each sequence, and use those in the intensity estimation. In our data set this thinning did not seem to change the intensity surface much and hence we retained the full eye movement sequences (see more discussion in Article III). In models introduced in this dissertation, the spatial heterogeneity term, $\alpha(\mathbf{x})$, is represented by the empirical intensity surface.

Dependency between consecutive fixations

The next component of the model should take into account dependency between fixations. Our idea is to control the length of the saccades, since the gaze has a tendency to make short correction saccades (Engbert et al., 2015) and long saccades are more rare. This feature is called *contextuality* in Article III. One way to control the effect of saccade lengths is to assume that the fixation $\mathbf{x}_{k+1}$ depends on the previous fixation $\mathbf{x}_k$, thus having the Markov property. This construction can be modelled by the kernel function $K(\mathbf{x}_k, \mathbf{x}_{k+1})$. Examples are

- truncated exponential kernel $K(\mathbf{x}_k, \mathbf{x}_{k+1}) \propto e^{-\lambda||\mathbf{x}_{k+1}-\mathbf{x}_k||}\, \mathbf{1}_W(\mathbf{x}_{k+1})$, $\lambda > 0$,
- truncated Gaussian kernel $K(\mathbf{x}_k, \mathbf{x}_{k+1}) \propto e^{-\frac{1}{2\sigma^2}||\mathbf{x}_{k+1}-\mathbf{x}_k||^2}\, \mathbf{1}_W(\mathbf{x}_{k+1})$, $\sigma > 0$,
- mixture kernel $K(\mathbf{x}_k, \mathbf{x}_{k+1}) = pK_1(\mathbf{x}_k, \mathbf{x}_{k+1}) + (1-p)K_2(\mathbf{x}_k, \mathbf{x}_{k+1})$, where $p \in [0, 1]$ is the probability of choosing the kernel K_1, and K_1 and K_2 are truncated kernel functions, for example, one can be the uniform distribution in W,
- separable kernel function $K(\mathbf{x}_k, \mathbf{x}_{k+1}) \propto g_1(||\mathbf{x}_{k+1} - \mathbf{x}_k||)g_2(\theta(\mathbf{x}_{k+1} - \mathbf{x}_k))\, \mathbf{1}_W(\mathbf{x}_{k+1})$, where g_1 is a model for saccade lengths and $g_2(\theta(\cdot))$ is a model for saccade angles θ.

Examples of the use of these kernels can be found in Article II (separable function) and in Article III (truncated Gaussian kernel).

Structured long-term dependence

If the process is suspected to change in time, the long-term dependence can be added to the model. In our case, it means that every time when the process takes a jump, the previous fixations of the process affect the probability law of the transition. The process is said to have *self-interaction* feature. In this dissertation, we consider two types of self-interaction: coverage-based and recurrence-based self-interaction. In practice, it means that the history of the process is included in the model by stochastic geometry, such as convex hull coverage or ball union coverage, or by the recurrence function introduced in Section 3.2. The purpose of the coverage-based self-interaction is to create a two-stage model where the target picture is first inspected globally and, after some time, the attention is allocated to finer details. The idea is that the function of the coverage of the fixations increases rapidly in the beginning

indicating the spread of the fixations throughout the window. This kind of behaviour is also supported by the concept of inhibition of return (IOR), which means that the attention towards previously visited elements of the picture is suppressed, but towards novel elements is encouraged (Posner and Cohen, 1984). Examples of such models are the *history-dependent rejection model* with coverage and the *history-adapted model* developed in Article III.

The recurrence-based self-interaction is based on the idea that after the initial view of the picture, the gaze will return to the interesting elements of the picture. We suggest to model this kind of long-term dependence in terms of the recurrence function. The process with recurrence-based self-interaction tends to favour locations close to previous fixations of the process. Such a behaviour causes spatial clustering, which is stronger than the clustering caused by the heterogeneous environment and short transitions. The history-dependent rejection model with recurrence is intended for modelling this kind of behaviour, see Article III and the spatio-temporal version of the model in Article IV.

The long-term dependence can be included into the model through the kernel function or as a separable component. If it is adapted to the kernel function, the kernel depends on the history, for instance, through a parameter, $K(\mathbf{x}_k, \mathbf{x}_{k+1}) = K_{\phi(\overrightarrow{\mathbf{x}}_k)}(\mathbf{x}_k, \mathbf{x}_{k+1})$. An example is the history-adapted model in Article III, where the information on the history is used to adjust the width of the Gaussian kernel. An alternative is to model long-term dependence as a separable component, denoted here as $\pi(\overrightarrow{\mathbf{x}}_{k+1})$, when the long-term dependence of the eye movement sequence is less confounded with the saccade lengths controlled by the kernel. The latter choice is used here and explained in detail below.

4.3 Maximum likelihood estimation for sequential spatial point processes

Finding the likelihood function for a spatial point process is often challenging and the maximum likelihood method can only be applied to certain point process model classes for which the likelihood is known, such as Poisson processes (Illian et al., 2008). In general, if the point process distribution can be constructed in terms of their probability densities, the maximum likelihood method can be applied. In some cases, the likelihood function is not known explicitly, but it can be approximated, for instance, by the pseudo-likelihood or by Monte Carlo methods (see e.g. Illian et al., 2008; Diggle, 2013, for more details).

The maximum likelihood method has been applied for finite sequential point processes, such as random sequential absorption model (van Lieshout, 2006c). In our case, the order of the events is known and the likelihood function for the developed sequential spatial point process models can be written up to a intractable normalizing constant (see Article III). For instance, the likelihood for the general sequential spatial point process model with all three components introduced above can be presented in form

$$L = f(\mathbf{x}_1) \prod_{k=1}^{n-1} c_k^{-1}\, \alpha(\mathbf{x}_{k+1})\, K(\mathbf{x}_k, \mathbf{x}_{k+1})\, \pi(\mathbf{x}_{k+1}, S(\overrightarrow{\mathbf{x}}_k, \mathbf{x}_{k+1})),$$

where

$\alpha(\cdot)$ describes spatial heterogeneity of the target,

$K(\cdot)$ is a Markovian kernel controlling spatial lengths of the transitions,

$\pi(\cdot)$ controls long-term dependence of the process, and

$S(\cdot)$ is a geometric feature of the previous points, such as convex hull coverage.

The long-term component can be thought to be a penalization term, such that when a new point is proposed according to the density proportional to $\alpha(\mathbf{x}_{k+1})K(\mathbf{x}_k, \mathbf{x}_{k+1})$, the term $\pi(\mathbf{x}_{k+1}, S(\overrightarrow{\mathbf{x}}_k, \mathbf{x}_{k+1}))$ decides whether the point is accepted or not. The terms c_k in the likelihood are normalizing integrals depending on the window W, on all the components introduced above and on the sequence $\overrightarrow{\mathbf{x}}_k$.

As an example of construction, we present the *history dependent rejection model* with the ball union coverage weighting (see Article III) which consists of the spatial heterogeneity term $\alpha(\mathbf{x}_{k+1})$, truncated Gaussian kernel $K(\mathbf{x}_k, \mathbf{x}_{k+1}) \propto e^{-\frac{1}{2\sigma^2}||\mathbf{x}_k - \mathbf{x}||^2} \mathbf{1}_W(\mathbf{x}_{k+1})$, and the ball union coverage as the long-term penalization term. Recall that the empirical intensity function represents here $\alpha(\mathbf{x})$. The penalization term depends on whether the new point is inside the ball union created by the previous points or not, that is,

$$\pi(\mathbf{x}_{k+1}, S(\overrightarrow{\mathbf{x}}_k, \mathbf{x}_{k+1})) = \begin{cases} \rho & \text{if } \mathbf{x}_{k+1} \in \mathrm{Bcov}(\overrightarrow{\mathbf{x}}_k) = \bigcup_{i=1}^{k} b(\mathbf{x}_i, r) \cap W \\ 1-\rho & \text{otherwise,} \end{cases}$$

where $\rho \in [0, 1]$. If $\rho > 0.5$, the process favours locations inside the ball union created by the previous points. The likelihood of the model can be written as

$$f_1(\mathbf{x}_1) \prod_{i=1}^{n-1} \frac{\alpha(\mathbf{x}_{i+1})\, e^{-\frac{1}{2\sigma^2}||\mathbf{x}_i - \mathbf{x}_{i+1}||^2} \left((1-\rho)\, \mathbf{1}_{W \setminus \mathrm{Bcov}(\overrightarrow{\mathbf{x}}_i)}(\mathbf{x}_{i+1}) + \rho\, \mathbf{1}_{\mathrm{Bcov}(\overrightarrow{\mathbf{x}}_i)}(\mathbf{x}_{i+1})\right)}{\int_W \alpha(\mathbf{u})\, e^{-\frac{1}{2\sigma^2}||\mathbf{x}_i - \mathbf{u}||^2} \left((1-\rho)\, \mathbf{1}_{W \setminus \mathrm{Bcov}(\overrightarrow{\mathbf{x}}_i)}(\mathbf{u}) + \rho\, \mathbf{1}_{\mathrm{Bcov}(\overrightarrow{\mathbf{x}}_i)}(\mathbf{u})\right) \mathrm{d}\mathbf{u}}, \quad (4.1)$$

where the density $f_1(\mathbf{x}_1)$ can be assumed to follow the empirical intensity function $\alpha(\mathbf{x})$, or modelling can be conditional on the first fixation $\mathbf{x}_1$. Note that here we fix the heterogeneity term $\alpha(\mathbf{x})$, but an option would be to use different heterogeneity term, for example, for the first 10 fixations than for the later ones. The normalizing integral depends on every parameter of the model and it has to be calculated for each step from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$, $k = 1, \ldots, n-1$. Hence the normalizing term is analytically intractable, but numerical integration methods, such as Riemann sums, can be used for its calculation.

Due to the way the sequential point process model is constructed here, the components of the model are dependent and the parameters of these components are confounded. The parameters of the model can be estimated one at a time, for instance, by the coordinate ascent method as is done in Article III, or simultaneously by applying optimisation algorithms after reparametrisation of the likelihood. In addition, one should consider whether to use all data in estimation. For example, if the long-term dependence of the model is based on recurrence penalization, this component does not have much effect in the beginning of the process, whilst the kernel function is dominating. In that case, it would be better to use the latter part of the eye movement sequence for estimating the parameters of the long-term component and use only the beginning of the sequence for estimating the kernel parameters. Another example is a process which spreads rapidly in the beginning due to the convex hull penalization. For this kind of model, the kernel starts to dominate after the sequence has covered most of the target window and hence the latter part of the sequence is informative for the kernel parameters.

4.4 Model diagnostics

Recall that our interest is in studying the uncertainty related to summary statistics (see Chapter 3) as well as in making group comparisons. After fitting a model for an eye movement sequence or sequences, we can apply Monte Carlo simulation, called also parametric bootstrap method in Efron and Tibshirani (1994, p. 53), to asses the uncertainty. The method is based on simulation of the fitted model with estimated parameters. First, we choose a summary statistic that describes the feature under study. In our case, the summaries we use are mainly functional in time, since we are interested in the fixation process. Second, we simulate realisations of the fitted model and estimate the chosen summary statistic for each realisation. These estimates can then be used to compute model-based envelopes which indicate statistical variation in that summary statistic under the model assumption. On the other hand, the envelope serves as a basis for simulation envelope test (Ripley, 1977), which give evidence against the null model if the data summary exceeds the envelope. This method has been traditionally used in model validation, which is reviewed next.

In spatial statistics the pointwise envelope is typically created by using the maximum and minimum value of the chosen functional summary statistics at each argument value (such as K-function at several distances r), calculated from the realisations of the model. Then, the data summary is plotted on the top of the envelope to check if the data summary exceeds it or not. However, the statistical significance of the simulated pointwise envelope is questionable, since with functional summary statistics we end up with the multiple testing problem when looking for deviations from the envelope at multiple points at the same time. This leads to invalid inference due to underestimation of the probability of type I error (significance level), as pointed out e.g. by Loosmore and Ford (2006). It is also a matter of number of simulations needed to obtain an envelope with a reasonable type I error probability (see Grabarnik et al., 2011). However, the pointwise envelope can be used for visualisation purposes and, as Baddeley et al. (2014) note, it can be plotted to assess what the result of the test would have been if we had chosen one particular argument value to perform the test, for example, if the distance r is fixed for the K-function prior simulations.

The difficulties in the pointwise envelope test can be bypassed by applying an alternative test procedure suggested by Loosmore and Ford (2006) relying on Barnard's Monte Carlo test (Barnard, 1963). The idea behind is following: the information in the data is suppressed to a single number, called the test statistic T_0. We generate m realisations of the null model and compute the value of the test statistic for each realisation. Assuming that the null hypothesis is true, the realisations and data are thought to be statistically equivalent. We can then compare the value T_0 to the simulated values $T_1, \ldots, T_m$, which gives us a p-value representing the significance of the observed result (more detailed explanation can be found e.g. in Baddeley et al., 2014). For functional summary statistics, one can use the Diggle-Cressie-Loosmore-Ford test (Baddeley et al., 2014),

$$T = \int_0^R (H(r) - H_{\text{theo}}(r))^2 \, dr,$$

where H is a summary statistic and H_{theo} is its theoretical value under the null hypothesis. Often H_{theo} is not known and it is calculated as an average of simulated and observed values of H (Baddeley et al., 2014). Another option for the test statistic would be the maximum deviation between H and its theoretical value (see e.g. Baddeley et al., 2014; Grabarnik et al.,

2011). Note that these kind of test statistics are sensitive to the upper value R. The problem with Monte Carlo tests arises, if the parameters of the null model are estimated from the data in question. This results that the Monte Carlo test is probably conservative (see e.g. Baddeley et al., 2014), since the true parameter value of the null model behind the data differs from the estimated parameter value which is used to generate the simulations.

Recently, several attempts have been made for creating envelopes that would truly represent certain significance level. Grabarnik et al. (2011) describe a way to calculate the type I error probability after performing the simulations, but they suggest to use deviation test for obtaining p-values and use envelopes just for visual purposes. Myllymäki et al. (2017) present two approaches, which allow to choose the significance level a priori and provide p-values and graphical representations as a result. These two global envelope tests are called the *rank envelope test* and the *scaled maximum absolute difference envelope test*. Both of these are robust to the choice of the upper value of the integral, but the first one requires a large number of simulations (e.g. 2499 if 5% significance level is pursued), whereas the latter one requires clearly less (99 for 5 % significance level) (Myllymäki et al., 2017). The global envelopes are applied in Article II, but only for graphical diagnostics.

Because of the complexity of eye movement sequences, finding one summary statistic describing the feature under study may be difficult or even impossible. We suggest to use several summaries at least for model validation. It should also be noted that the summary characteristics or statistics used for model validation should be chosen to be different in nature than the one or ones used in fitting the model (Illian et al., 2008) or in model construction. In the articles included in this dissertation, we have used the following summary statistics in model validation: convex hull and ball union coverage, scanpath length, recurrence function and transition probabilities between four specified areas of the target space. In addition, the intensity surface is used in checking the goodness of fit of the models, but only for visualising purposes. The summary characteristics of point process models, such as the K-function, can also be used for evaluation of the models. Note that we use here pointwise envelopes as graphical diagnostics in model validation, which is allowed since we do not use them for statistical testing.

Chapter 5

Decomposition of art viewing: an example

In this chapter, the methods introduced above are applied to eye movement data. The idea is to experiment, whether long-term dependence exists in an eye movement sequence measured from one subject. We concentrate on the first 100 fixations occurred when the subject was looking at the painting *La croisée des destins* by Risto Suomi (see, Figure 1.1). These 100 fixations were measured during the first 41 seconds of inspection. On the painting there are some clearly salient elements, namely the shark and the rabbit, which will most probably gain attention. It can be expected that these elements will be looked at quite early and, as the saccades in Figure 1.1 indicate, the gaze jumps between these elements. The aim of this example is to study, if we can decompose this eye movement sequence using the new sequential spatial point process modelling. Is the gaze jumping around on the painting randomly, or is there a learning effect? The sequence of fixations during the first minute of inspection on the painting can be seen in Figure 1.1 and the convex hull and ball union coverages of the eye movements during the first 10 seconds are illustrated in Figure 3.1, left. Indeed, these figures indicate that the subject has looked at the most salient elements in the beginning of the inspection.

In order to study the long-term dependence, we decompose the fixation sequence into three components described in Section 4.2.2. In particular, we fit the history-dependent rejection model with ball union coverage (4.1) to the data. The ball union coverage is chosen to be used in the penalization term, because the target painting includes a few visually very salient elements that will probably attract attention. According to this model, a new point outside the ball union, created by the previous points, is accepted with probability $1 - \rho$ and inside the ball union with probability ρ. If $\rho = 0.5$, the model coincides with the random walk in heterogeneous space. To study the long-term dependence, we fit both the random walk model and the history-dependent rejection model to the data.

The spatial heterogeneity term $\alpha(\mathbf{x})$ and the ball radius $r = 50$ pixels are fixed before model fitting. The radius is chosen such that one ball located on the top of the shark or the rabbit covers the whole object, since we assume that these salient objects can be recognised with one fixation. As the heterogeneity term $\alpha(\mathbf{x})$ we use the empirical intensity that is estimated from all fixations in the data set related to the painting of Risto Suomi and measured from 19 participants, excluding the fixations of the subject under study. This kind of pooled empirical intensity surface serves as the empirical saliency map (see Section 1.4). The size of the balls

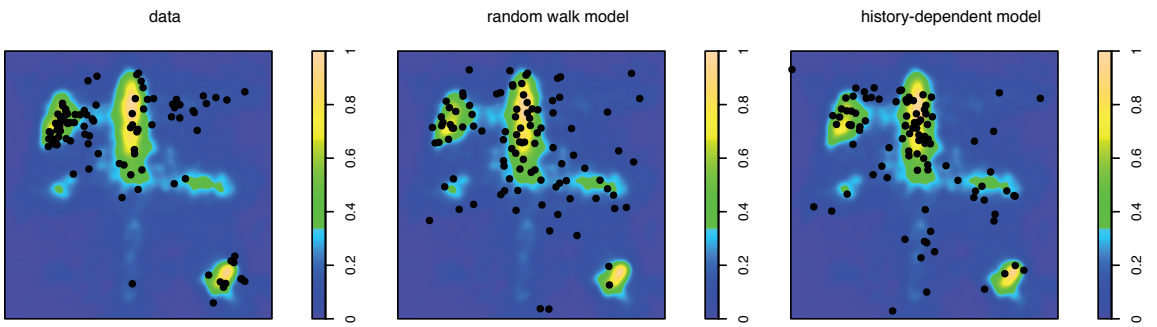

Figure 5.1: Point maps of the data (left), a realisation of the random walk model (middle) and a realisation of the history-dependent rejection model (right) with the estimated intensity surface (the empirical saliency map).

used in the penalization term with respect to the size of the painting is visualised in Figure 3.1, left. The coordinate ascent algorithm is used for fitting the models to the fixation sequence.

For the random walk model, we obtain $\hat{\sigma} = 240$ as an estimate of the kernel parameter, whereas the parameter estimates of the fitted history-dependent rejection model are $\hat{\sigma} = 260$ and $\hat{\rho} = 0.8$. In both cases, the kernel is quite flat and allows the process to make rather long jumps according to the estimates of the kernel parameter. In terms of long-term dependence, the two fitted models differ. The estimate $\hat{\rho} = 0.8$ of the history-dependent rejection model indicates that the locations inside the ball union coverage are favoured compared to those outside the coverage.

Next, we simulate realisations of these fitted models. The first 100 fixations of the subject on the estimated intensity surface is presented in Figure 5.1, left. The realisations of the fitted random walk model and of the fitted history-dependent rejection model are presented in the middle and on the right hand side of Figure 5.1. The point map resulting from the random walk model is less clustered compared to the data indicating that the spatial heterogeneity (estimated from the pooled data) does not alone explain the locations of fixations. The point map of the realisation of the history-dependent rejection model performs better in terms of clustering, but the points seem to be allocated somehow differently than the data. This is the drawback of estimating the spatial heterogeneity from a set of eye movement sequences: variation between subjects is large and when the information is summarized over several subjects we lose subject-level information.

In order to check the goodness-of-fit of these two models, we simulate 99 realisations of each fitted model. From these simulations we calculate the pointwise envelopes for the chosen summary statistics. We suggest to use several summary statistics in model evaluation, because none of the summaries alone are able to describe all the important features of the process. Here, we employ the data summaries represented in Chapter 3 together with the inhomogeneous K-function (see equation (2.2)). The same empirical intensity surface is used for $\lambda(\mathbf{x})$ when estimating K_I as we used for $\alpha(\mathbf{x})$. The results can be seen in Figures 5.2, 5.3 and 5.4.

For the random walk model, the empirical ball union coverage with radius $r = 35$ exceeds the pointwise lower envelope and the empirical cumulative recurrence function exceeds the pointwise upper envelope, see Figure 5.2. This means that the model favours locations that

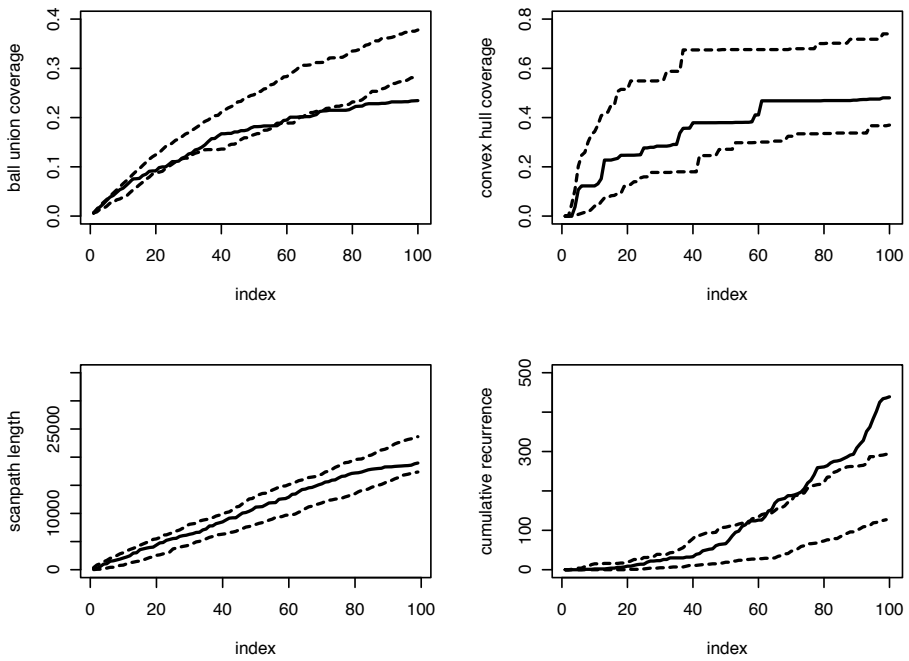

Figure 5.2: Ball union coverage with $r = 35$ (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with $r = 50$ (bottom right) for the data. The dashed lines represent the simulated pointwise envelopes based on 99 realisations of the fitted random walk model ($\hat{\sigma} = 240$).

are outside the ball union coverage and are hence apart from the previous points. Since we used the ball union coverage with $r = 50$ in model construction, here we used the smaller radius of size $r = 35$. Based on these summary statistics, the random walk model with spatial heterogeneity and Markovian kernel components does not fit the data well. In the case of the history-dependent rejection model, all the four data summaries as well as $\hat{K}_I$-function remain inside the pointwise envelopes, see Figures 5.3 and 5.4. Note, however, that the inhomogeneous K-function was not able to distinguish the random walk from the history-dependent rejection model because it does not take the time into account.

To conclude, the history-dependent rejection model seems to fit well to the fixation sequence under study. The estimated model parameter $\hat{\rho} = 0.8$ indicates that the eye movement sequence is not just a random walk, but the locations inside the ball union coverage created by the previous points are favoured. From Figure 3.1 one can deduce that the ball union coverage covers the most salient elements, the shark and the rabbit, during the first 10 seconds of inspection. After that, the increase in the ball union coverage curve declines meaning that the gaze does not jump far away from the previous fixations. This kind of behaviour can be interpreted as an evidence of long-term dependence: the process tends to favour the already visited locations.

In this empirical example, we were able to find indications of learning in the sequence with a rather parsimonious model by using information only on the order of fixations but not on their occurrence times.

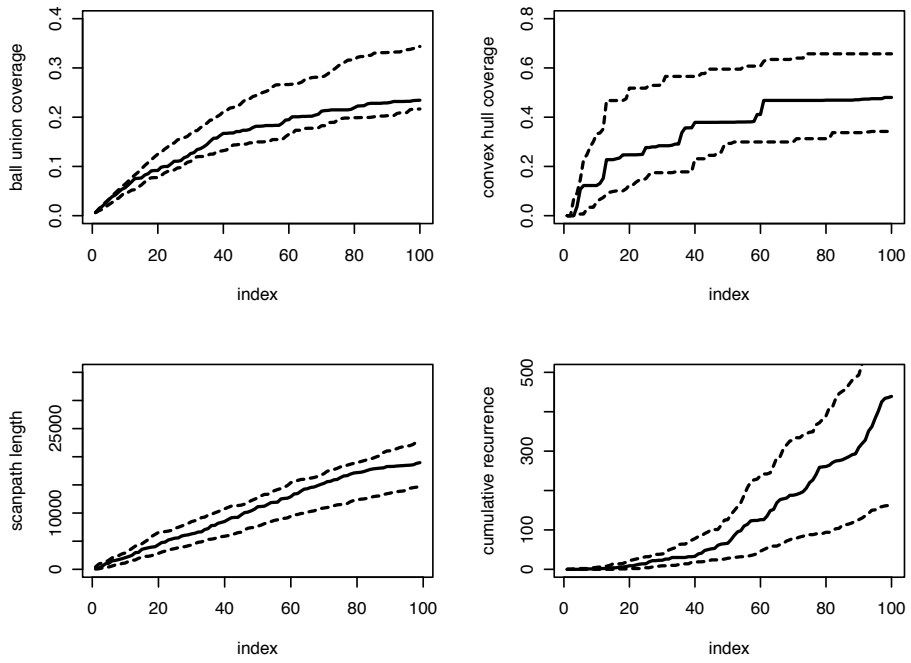

Figure 5.3: Ball union coverage with $r = 35$ (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with $r = 50$ (bottom right) for the data. The dashed lines represent the simulated pointwise envelopes based on 99 realisations of the fitted history-dependent rejection model ($\hat{\sigma} = 260, \hat{\rho} = 0.8$).

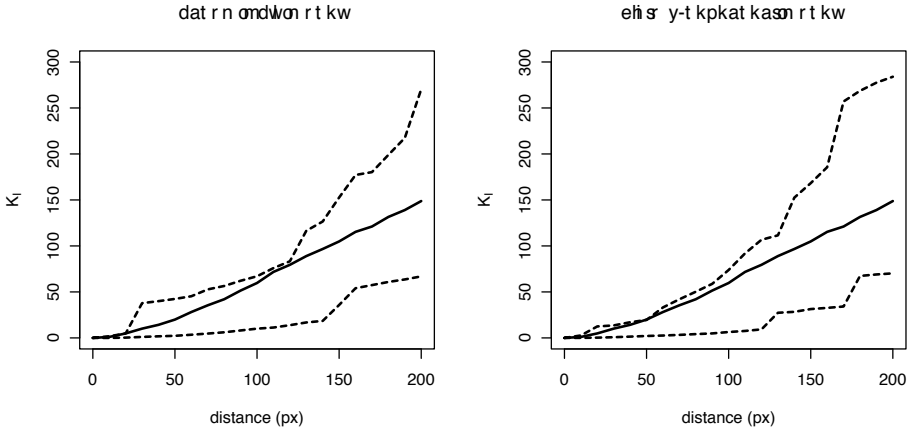

Figure 5.4: $\hat{K}_I$-function for the data with the simulated pointwise envelope based on 99 realisations of the fitted random walk model (left) or of the history-dependent rejection model (right).

Chapter 6

Discussion

This dissertation aimed at contributing to the eye movement methodology, where data consist of movements of the gaze recorded with an eye tracker. The research questions studied here arose from an art study and the associated experimental data set. The idea behind the experiment was to compare visual processing of novices and non-novices in art viewing, and therefore we developed methods that can be used for group comparisons. In order to understand visual processing of paintings, we studied how the eye movement process evolves in time. This aspect is often neglected in the analyses due to complexity of the data or lack of availability of statistical methods. Here, the data set allowed us to study both the dynamics and long-term dependence in eye movement sequences due to the sufficiently long inspection time.

In this pioneering work, parallel with two other studies using point process statistics for eye movement data (see Barthelmé et al., 2013; Engbert et al., 2015), we studied how spatial and spatio-temporal point process statistics and tools therein can be applied to eye movement data. Instead of jumping right into spatio-temporal point process framework, we started from spatial point process statistics and design-based approaches and then proceeded to sequential spatial point processes by taking into account the temporal order of fixations. Finally, we included also the occurrence times and presented a spatio-temporal point process model for fixations. This step by step approach helped us to understand the limitations of the marginal processes of a spatio-temporal point process. From a statistical perspective, we enlarged the field of spatio-temporal point process statistics by introducing new summary statistics, by presenting useful and interpretable point process models, and by applying likelihood-based inference.

The novelty of our work in the eye movement context is in the study of the evolution of the eye movement process instead of concentrating just on the static pattern of fixations or other strongly aggregated data summaries. We took into account the temporal nature of the eye movement process both through functional summary statistics and by considering the history of the process in modelling. The functional summary statistics that we proposed for describing the temporal evolution of the process can also be used in model construction and evaluation. In addition, we presented subject level models for decomposing features, such as long-term dependence, from an eye movement sequence. These models are parsimonious sequential point process models, but their usefulness was illustrated in this work.

The sequential spatial point process models introduced here could be further developed, for instance, by modifying the penalization term or by letting the spatial heterogeneity term

vary in time. However, the estimation of spatial heterogeneity is problematic as is discussed in Article III, in Barthelmé et al. (2013) and in Engbert et al. (2015), since the dependency between consecutive fixations is a cause of extra clustering. Another issue related to the sequential point process models is the computation of the normalizing integrals in the likelihood. Our choice to use numerical integration (Riemann sums) is rather slow to compute for long eye movement sequences, because the integral has to be calculated for every step of the process. Computation of the integral and the likelihood itself could most likely be improved by other approximative methods.

We brought eye movement data to a spatio-temporal point process context with the idea that we start from the sequential point process modelling of fixation locations being the most informative part of eye movement sequence. The spatio-temporal point process model is then constructed under a condition that the fixation locations conditional on the history of the process do not depend on the occurrence times of the previous fixations. This condition, being a generalisation of the space-time separability, is not a severe restriction in this context. The occurrence times of fixations are modelled conditional on the locations. One of the attempts for the future would be to study further the relationship between fixation locations and occurrence times. This information would help in building more general spatio-temporal point process models. On the other hand, the use of spatio-temporal point process methods requires reasonable amount of events, which means that the experiment and the resulting eye movement sequence should be long enough, or replicates should be available.

Altogether, this dissertation opens new views to eye movement methodology. The growing use of eye tracking method is constantly raising new research questions and thus demands more sophisticated tools for the analysis. We believe that the point process statistics helps to deal with these kind of complex data, if only the researchers would be aware of the methods and tools were available. The tools introduced here, such as data summaries and models, are rather easy to implement, whereas K-functions and other point process characteristics can be already found in R (R Core Team, 2016), see e.g. packages `spatstat` (Baddeley et al., 2015) and `stpp` (Gabriel et al., 2013). The work presented in this dissertation gives good premises for developing the methodology further.

Summary of original publications

Article 1: Spatial point process statistics in understanding eye movement data

Eye movement sequence can be considered as a realisation of a spatio-temporal point process. However, before modelling such complex data, more statistical analysis methods are needed for understanding the nature of the eye movement process. Here, we consider simple model-free methods that are based on marginals of the underlying spatio-temporal point process. More precisely, we apply spatial point process statistics for detecting changes in the behaviour of a fixation sequence in terms of spatial clustering and fixation durations. We compare the early fixations of an eye movement sequence with the late fixations of the same sequence by using Ripley's K-function. Since the original K-function assumes stationarity, which is not fulfilled by our data, we apply the case-control method and study the difference of two K-functions. Another way to bypass this problem is to apply the inhomogeneous K-function, for which the first-order intensity has to be estimated. We estimate the intensity using the pooled eye movement data of other subjects participating the experiment and study also the difference of the inhomogeneous K-functions of the two groups. This comparison reveals extra clustering of fixations when the features of the painting are taken into account. Finally, we investigate the stability of mark-correlation in an eye movement sequence by adding the durations of fixations as marks.

Article 2: What we look at in paintings: A comparison between experienced and inexperienced art viewers

This article introduces statistical tools for analysing eye movements in two groups. We concentrate on comparing the eye movements of novice and non-novice art viewers on one painting called Koli landscape, but we also report the results for the five other paintings in our data set. In addition, we study the consistency of the distribution of fixation durations between different paintings in order to find out how the painting itself affects the eye movements. The eye movement process on Koli landscape painting is illustrated with several static and functional summary statistics, which are rather novel in eye movement methodology, describing spatial nature and temporal dynamics of the process. To have a more detailed comparison of the two groups, we introduce a group-level spatio-temporal point process model for fixations. The model is rather simple, but is able to catch the variation related to the functional summary statistics reasonably well. However, the developed model does not catch the full temporal dynamics shown in the fixation process, because it does not consider the long-term history of the process, only the previous location.

Article 3: Deducing self-interaction in eye movement data using sequential spatial point processes

The eye movement process can be thought to contain learning due to the visual information obtained from the target and cognitive mechanisms affecting the movement of the gaze. The objective of this article is to introduce sequential spatial point process models for detecting such dynamic structures in eye movement data. What is special for these models is that they take the history of the process into account through a so-called self-interaction component. The self-interaction is created by two alternative mechanisms: coverage or recurrence. The first one can be used to model global scanning in the beginning of the process, while the latter can model the behaviour where the eye movement process tends to return to already visited areas. Spatial heterogeneity, that is usually present in eye movement data, is controlled by estimating the pooled intensity of the eye movements of other participants of the experiment considered as replicates. The models also control the lengths of the saccades by the Gaussian kernel. An empirical example illustrating the use of one of the developed model is given revealing a large variation between the individuals.

Article 4: Detecting the long-memory effect in an eye movement sequence

This paper is motivated by the idea that the gaze has a tendency to return to the pre-examined areas of the target, which we call here as the long-memory effect. This kind of long-term interaction causes spatial clustering in a long time range, which differs from the traditionally studied space-time clustering. The aim is to experiment whether such behaviour could be detected statistically. Therefore, we expand the sequential spatial point process model introduced in Article 3 to spatio-temporal framework by adding the temporal model for the occurrence times. By applying the resulting model, we study how the second-order characteristics of spatio-temporal point processes, namely the pair correlation function and the K-function, can be employed for detecting the long-memory effect in the sequence. A challenge is the spatial heterogeneity that have to be taken into account both in modelling and in applying the summaries. The ability of the suggested method in detecting the effect is illustrated by synthetic data.

Bibliography

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

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.

Baddeley, A. J., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R.* Chapman and Hall/CRC Press, Boca Raton.

Barnard, G. A. (1963). Discussion on paper by M. S. Bartlett. *Journal of Royal Statistical Society: Series B*, 25:294.

Barthelmé, S., Trukenbrod, H., Engbert, R., and Wichmann, F. (2013). Modelling fixation locations using spatial point processes. *Journal of Vision*, 13(12):1–34.

Borji, A. and Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207.

Buswell, G. T. (1935). *How People Look at Pictures - A Study of the Psychology of Perception in Art.* The University of Chicago Press, Chicago, Illinois.

Castelhano, M. S. and Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 33(4):753–763.

Castelhano, M. S. and Henderson, J. M. (2008). Stable individual differences across images in human saccadic eye movements. *Canadian Journal of Experimental Psychology*, 62(1):1–14.

Chiu, S. K., Stoyan, D., Kendall, W. S., and Mecke, J. (2013). *Stochastic Geometry and its Applications, 3rd ed.* Wiley, Chichester.

Cuzick, J. and Edwards, R. (1990). Spatial clustering for inhomogeneous populations (with Discussion). *Journal of the Royal Statistical Society: Series B*, 52:73–104.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods.* Springer-Verlag, New York.

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure.* Springer-Verlag, New York.

Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns.* Chapman and Hall/CRC Press, Boca Raton.

Diggle, P. J., Besag, J., and Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, pp. 659–667.

Diggle, P. J. and Chetwynd, A. G. (1991). Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics*, 47:1155–1163.

Diggle, P. J., Chetwynd, A. G., Häggkvist, R., and Morris, S. (1995). Second-order analysis of space-time clustering. *Statistical Methods in Medical Research*, 4:124–136.

Diggle, P. J., Gómez-Rubio, V., Brown, P. E., Chetwynd, A. G., and Gooding, S. (2007). Second-order analysis of inhomogeneous spatial point processes using case-control data. *Biometrics*, 63(2):550–557.

Diggle, P. J., Kaimi, I., and Abellana, R. (2010). Partial likelihood analysis of spatio-temporal point process data. *Biometrics*, 66(2):347–354.

Doksum, K. A. and Sievers, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika*, 63:421–434.

Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments & Computers (BRMIC)*, 34:455–470.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap.* Chapman and Hall/CRC Press.

Engbert, R., Trukenbrod, H., Barthelmé, S., and Wichmann, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *Journal of Vision*, 15(1):1–17.

Evans, J. V. (1993). Random and cooperative sequential adsorption. *Reviews of modern physics*, 65(4):1281.

Foulsham, T. and Underwood, G. (2008). Whan can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 8(2):6–6.

Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63(1):43–51.

Gabriel, E., Rowlingson, B., and Diggle, P. J. (2013). `stpp`: An `R` package for plotting, simulating and analysing spatio-temporal point patterns. *Journal of Statistical Software*, 53(2):1–29.

Goldberg, H. J. and Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24(6):631–645.

González, J. A., Rodríguez-Cortés, F. J., Cronie, O., and Mateu, J. (2016). Spatio-temporal point process statistics: A review. *Spatial Statistics*, 18:505–544.

Grabarnik, P., Myllymäki, M., and Stoyan, D. (2011). Correct testing of mark independence for marked point patterns. *Ecological Modelling*, 222(23):3888–3894.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures.* OUP Oxford.

Iftimi, A., Martínez-Ruiz, F., Santiyán, A. M., and Montes, F. (2015). Spatio-temporal cluster detection of chickenpox in Valencia, Spain in the period 2008–2012. *Geospatial Health*, 10:54–62.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns.* Wiley, Chichester.

Itti, L., Koch, C., and Nieburg, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1254–1259.

Jarodzka, H., Holmqvist, K., and Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 symposium on eye-tracking research & applications.*, pp. 211–218. ACM.

Jensen, E. B. V., Jónsdóttir, K. Ý., Schmiegel, J., and Barndorff-Nielsen, O. E. (2007). Spatio-temporal modelling – with a view to biological growth. In Finkenstädt, B., Held, L., and Isham, V. (Eds.), *Statistical methods for spatio-temporal systems, Monographs on statistics and applied probability*, pp. 47–75. Chapman & Hall/CRC, Boca Raton.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87:329–355.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-U., Lee, M.-H., Chiou, G.-L., Liang, J.-C., and Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10:90–115.

van Lieshout, M. N. M. (2006a). Campbell and moment measures for finite sequential spatial processes. In Hušková, M. and Janžura, M. (Eds.), *Proceedings Prague Stochastics 2006*, pp. 215–224, Prague. Matfyzpress.

van Lieshout, M. N. M. (2006b). Markovianity in space and time. *IMS Lecture Notes—Monograph Series*, 48:154–168.

van Lieshout, M. N. M. (2006c). Maximum likelihood estimation for random sequantial adsorption. *Advances in Applied Probability (SGSA)*, 38:889–898.

Locher, P. J. (2006). The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. *Psychological Science*, 48:106–114.

Locher, P. J., Gray, S., and Nodine, C. F. (1996). The structural framework of pictorial balance. *Perception*, 25(12):1419–1436.

Locher, P. J., Krupinski, E. A., Mello--Thoms, C., and Nodine, C. F. (2007). Visual interest in pictorial art during an aesthetic experience. *Spatial Vision*, 21:55–77.

Loosmore, N. B. and Ford, E. D. (2006). Statistical inference using the G or K point pattern spatial statistics. *Ecology*, 87:1925–1931.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC Press, Boca Raton.

Myllymäki, M., Mrkvička, T., Grabarnik, P., Seijo, H., and Hahn, U. (2017). Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B*, 79(2):381–404.

Noton, D. and Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11(9):929–942.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, 50(2):379–402.

Parkhurst, D., Law, K., and Nieburg, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42:107–123.

Penttinen, A. and Stoyan, D. (1989). Statistical analysis for a class of line segment processes. *Scandinavian Journal of Statistics*, 16:152–168.

Posner, M. I. and Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, 32:551–556.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62:1457–1506.

Ripley, B. D. (1977). Modelling spatial patterns (with Discussion). *Journal of the Royal Statistical Society: Series B*, 39:172–212.

Salvucci, D. D. and Anderson, J. R. (1998). Tracing eye movement protocols with cognitive process models. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, pp. 923–928, Hillsdale, NJ. Lawrence Erlbaum Associates.

Salvucci, D. D. and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pp. 72–78. ACM.

Talbot, J., Tarjus, G., Van Tassel, P. R., and Viot, P. (2000). From car parking to protein adsorption: An overview of sequential adsorption processes. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, 165(1):287–324.

Tatler, B. W., Baddeley, R. J., and Gilchrist, I. D. (2005). Visual correlates of fixation selection: Effects of scale and time. *Vision Research*, 45:643–659.

Tatler, B. W., Baddeley, R. J., and Vincent, B. T. (2006). The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, 46(12):1857–1862.

Underwood, G., Humphrey, K., and Foulsham, T. (2008). Knowledge-based patterns of remembering: Eye movement scanpaths reflect domain experience. In Holzinger, A. (Ed.), *HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008*, pp. 125–144. Springer Berlin Heidelberg.

Yarbus, A. L. (1967). *Eye Movements and Vision.* Plenum Press, New York.

II

What we look at in paintings: A comparison between experienced and inexperienced art viewers

Ylitalo, A.-K., Särkkä, A. and Guttorp, P.

The Annals of Applied Statistics, 10(2): 549–574, 2016.

doi:10.1214/16-AOAS921

The Annals of Applied Statistics
2016, Vol. 10, No. 2, 549–574
DOI: 10.1214/16-AOAS921

WHAT WE LOOK AT IN PAINTINGS: A COMPARISON BETWEEN EXPERIENCED AND INEXPERIENCED ART VIEWERS

BY ANNA-KAISA YLITALO[*,1], AILA SÄRKKÄ[†,2] AND PETER GUTTORP[‡,§]

University of Jyvaskyla[*],
Chalmers University of Technology and University of Gothenburg[†],
University of Washington[‡] and Norwegian Computing Center[§]

How do people look at art? Are there any differences between how experienced and inexperienced art viewers look at a painting? We approach these questions by analyzing and modeling eye movement data from a cognitive art research experiment, where the eye movements of twenty test subjects, ten experienced and ten inexperienced art viewers, were recorded while they were looking at paintings.

Eye movements consist of stops of the gaze as well as jumps between the stops. Hence, the observed gaze stop locations can be thought of as a spatial point pattern, which can be modeled by a spatio-temporal point process. We introduce some statistical tools to analyze the spatio-temporal eye movement data, and compare the eye movements of experienced and inexperienced art viewers. In addition, we develop a stochastic model, which is rather simple but fits quite well to the eye movement data, to further investigate the differences between the two groups through functional summary statistics.

1. Introduction. Eye movements are outcomes of cognitive processes in the human brain, and can be recorded with high spatial and temporal resolution by computerized eye trackers. Eye movements provide valuable information about cognitive processes [Duchowski (2002), Rayner (1998, 2009)], and tracking of them is analyzed in a range of different areas, such as language and music reading [Kinsler and Carpenter (1995), Rayner (1998)], psychology [Findlay (2009)], and marketing research [Nagasawa, Yim and Hongo (2005)].

The first measurements of eye movements were made in 1879 independently by M. Lamare in France and Ewald Hering in Germany [Wade (2010)]. Both researchers used an acoustic approach, and noticed, much to everyone's surprise, that in reading text the gaze moves in jerks between points of rest. These extremely rapid movements are called *saccades*. They are essentially involuntary, so once a saccade starts it cannot be interrupted [Findlay (2009)]. Thus, researchers can predict saccade lengths from very few observations early in the saccade [Komogartsev, Ryu and Koh (2009)]. The points of rest, periods in which the gaze

Received October 2014; revised December 2015.
[1]Supported in part by the Finnish Doctoral Programme in Stochastics and Statistics and by the Academy of Finland (Project number 275929).
[2]Supported by the Knut and Alice Wallenberg Foundation.
Key words and phrases. Coverage, intensity, point process, shift function, transition probability.

is staying relatively still around a location of the target space, are called *fixations* [Barlow (1952)]. Eye movements can thus be represented as an alternating sequence of fixations and saccades. In this paper we will mainly focus on sequences of fixation locations and times in the target space, and call such a sequence of observations a fixation process. Our targets are pictures of paintings and the data consist of recorded eye movements of subjects on the paintings. We will introduce some new statistical tools and a model to analyze the fixation process by means of group comparisons.

The fixation process is regarded as a spatio-temporal point process. We are aware of only few studies that use point process methodology to analyze fixation processes. Barthelmé et al. (2013) use inhomogeneous Poisson processes to model fixation locations, and Engbert et al. (2015) perform some preliminary analysis of clustering of the fixations by using inhomogeneous pair correlation function. Furthermore, they construct a dynamic model for saccade selection by discretizing the space.

We are interested in studying how people look at art. The first such study was carried out using two movie cameras by Buswell (1935). His findings indicated that much of the theory of how people look at art needed reformulation. Another interesting eye movement study related to arts is brought out by Locher (2006). He developed a two-stage model for art viewing, where the first stage of viewing of a painting was to obtain a general view of its structure and semantics. In Locher's study, subjects describing a painting orally while viewing it started to describe the painting only after a couple of seconds after they had begun to look at it. At the second stage of the model the viewer focuses on some interesting features, which are analyzed from an aesthetic point of view.

In this paper, we concentrate on comparing the fixation process of inexperienced and experienced art viewers, *novices* and *non-novices* for short. Already in his pioneering work Buswell (1935) compared the fixation durations of 61 non-novices and 117 novices (originally "art experts" and "lay persons") by comparing the group averages based on the 25 first fixations. He repeated the experiment for seven paintings, and in each case he concluded that the non-novices made shorter fixations than the novices. Nevertheless, when comparing the fixations in the areas of the subdivided picture, he did not report any major differences between the two groups. However, there are some studies where some differences have been found. In Kristjanson and Antes (1989) ten paintings were shown to a group of non-novices and a group of novices ("artists" and "non-artists" in the original paper). All the subjects were familiar with three of the paintings (same for all subjects), and the non-novices group was familiar with another three paintings, while the remaining paintings were unknown to all subjects. The only statistically significant difference between the non-novice and novice groups found was that non-novices made more fixations on the parts of the paintings that were not centers of interest when viewing unfamiliar paintings compared to novices. The authors concluded that the non-novices made a more thorough investigation of all areas

of the unfamiliar paintings than the novices. The final conclusion was that it is conceivable that the non-novices extract a different amount or different kind of information during each fixation, even though the pattern of fixations may vary little from the pattern of novices.

Vogt and Magnussen (2007) used eye movement patterns to investigate how non-novices and novices ("artists" and "artistically untrained people" in the original paper) view art. Participants were shown realistic and abstract works of art under two conditions: one asking them to free scan the paintings, and the other asking them to memorize them. Participants' eye movements were tracked as they either looked at the images or tried to memorize them, and their recall for the memorized images was recorded. The researchers found no differences in the fixation frequency or duration between picture types for non-novices and novices. However, across the two conditions, the novices had more short fixations while free scanning the works, and fewer long fixations while trying to memorize. Non-novices followed the opposite pattern. In addition, non-novices spend more time than novices by scanning the areas which were not defined as regions of interest. There was no statistically significant difference in the recall of the images across groups, except that non-novices recalled abstract images better than novices and remembered more pictorial details.

Here, we investigate whether we can find any differences between how non-novices and novices view a particular painting by using a new set of tools to study eye movements. As mentioned earlier, the fixation process is described by a spatio-temporal point process, and intensities of the point processes and distributions of the fixation and saccade durations are used to compare the eye movements of novices and non-novices. We are mainly concentrating on the eye movements of 20 subjects, 10 novices and 10 non-novices, on the painting *Koli landscape* by Eero Järnefelt [Sundell (1986)] but give some results based on the other five paintings as well. For example, we investigate whether the fixation duration distributions of the novice and non-novice groups remain the same from painting to painting. Finally, based on the data analysis, we construct a rather simple stochastic model that further helps to describe the fixation process and to investigate differences between the non-novices and novices in terms of functional summary statistics.

The paper is organized as follows. The experimental setup and the data from the cognitive art research experiment as well as the results from the preliminary data analysis are described in Section 2. In Section 3, we show the results of a comparison of the two groups, novices and non-novices. A stochastic model is constructed in Section 4, and it is used to make further comparisons of the groups for the Järnefelt painting in Section 5. Finally, in Section 6 we discuss the results.

2. Data and data analysis.

2.1. *Art experiment.* Twenty test subjects participated in an experiment where the task was to observe six pictures of paintings on a computer screen, each during three minutes, and describe orally the mood in each painting while their eye

movements and voice were recorded. The paintings can be seen in Figure 1. The eye movements were recorded by the SMI iView X™ Hi-Speed eye tracker with a temporal resolution of 500 Hz. The resolution of the target screen was 1024×768 pixels and the distance between a participant's head and the screen was around 85 cm. A forehead rest was used in order to avoid redundant movement of the head. The data were collected by Pertti Saariluoma and Sari Kuuva (University of Jyväskylä) and María Álvarez Gil (University of Salamanca) with technical help from Jarkko Hautala and Tuomo Kujala (University of Jyväskylä). All subjects were students at the University of Jyväskylä at the time of the study. Ten of the subjects were either art students (8) or students who had studied art history and frequently visited art exhibitions (2). The remaining ten subjects were students who did not have art as their major or their hobby. We call the participants in the first group *non-novices* and participants in the latter group *novices*. Five of the participants were men (three novices and two non-novices) and 15 women (seven novices and eight non-novices).

In this article our main purpose is to introduce some new tools for eye movement analysis. Therefore, instead of analyzing all the six paintings in detail, we concentrate mainly on the eye movements on one of the paintings, namely, the Järnefelt painting Koli landscape, shown in Figure 1(a). The resolution of the image of the painting is 770×768 pixels, hence, it does not fill the whole computer screen and there are white areas on both sides of the painting. For some subjects, some of the fixations were located in these white areas outside the painting (Figure 2, right) and excluded from the analysis. We will treat saccades going outside the image as missing values.

2.2. *Data description*. We first look at the eye movement data as a whole and later (Section 3) compare the eye movements of novices and non-novices. An important way of describing the viewing of the painting is to look at a smoothed plot of all fixation locations of all subjects, which represents the viewing foci of the painting. Technically, when describing the locations of fixations by a spatial point process this is an estimate of the intensity of the fixation process, marginalized over time. We use the R package `spatstat` for the intensity estimation [Baddeley, Rubak and Turner (2015)]. A chi-square test for quadrat counts clearly rejects the hypothesis of constant intensity ($p < 0.001$). The areas of the painting that have been visited most by the subjects are shown in Figure 3 (left). We can see that the trunk of the main tree going down almost the entire length of the painting and the top of the small tree next to it seem to be the areas of particular interest in the painting. In Figure 3 (right) one can see the least visited areas, the edges of the painting in this case.

The fixation process varies a lot between subjects, and hence it is interesting to look at some summary statistics for the data. The supplemental article by Ylitalo, Särkkä and Guttorp (2016a) includes a table which shows the total number of fixations and the number of short fixations (less than 40 ms) inside the picture with

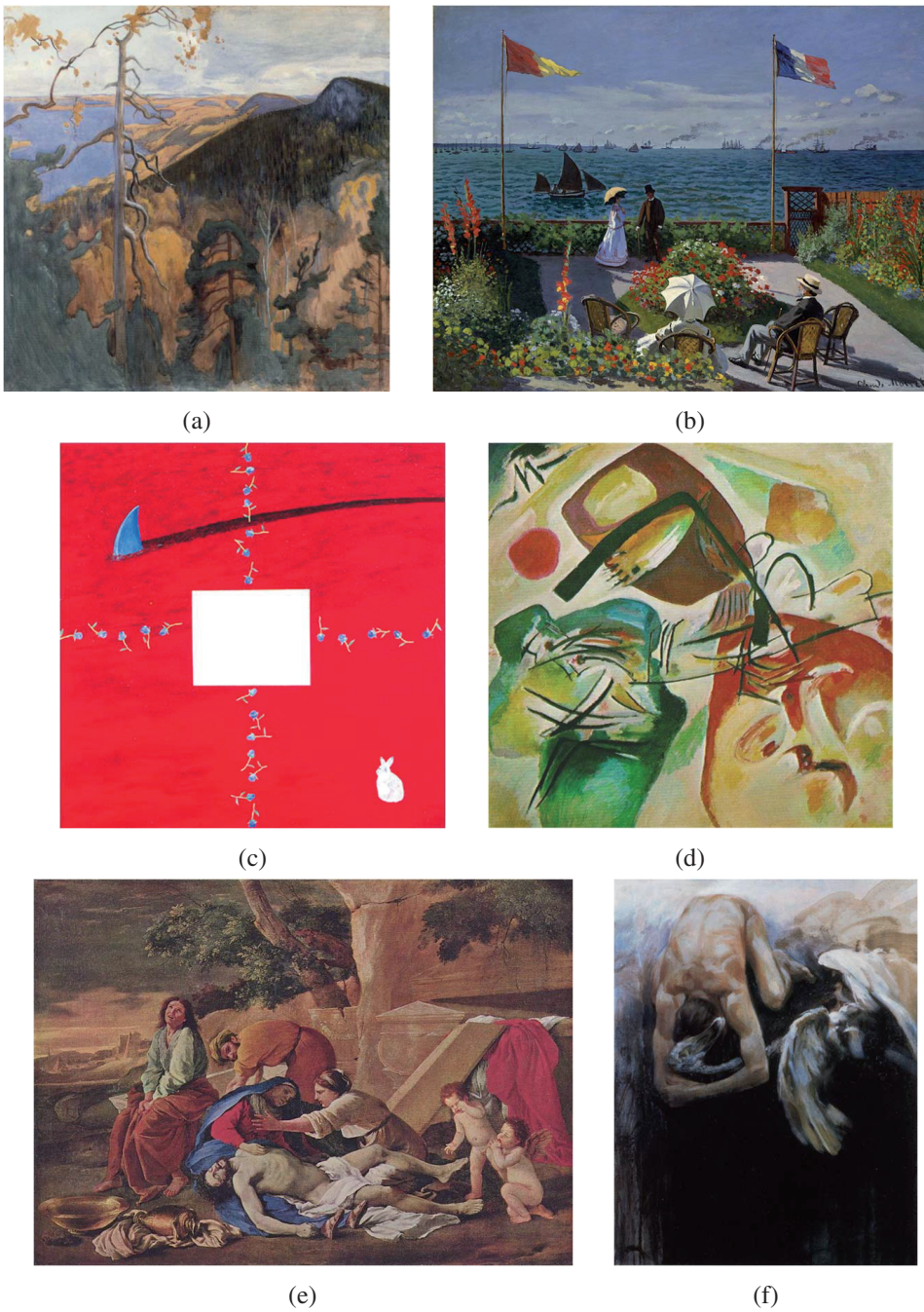

FIG. 1. *The stimulus paintings used in the experiment*. (a) *Eero Järnefelt—Koli-landscape* (*the turn of* 19*th and* 20*th century*), *source*: *Sundell* (1986)*/public domain*. (b) *Claude Monet—Terrasse à Sainte-Adresse* (1867), *source*: *The Metropolitan Museum of Art/public domain*. (c) *Risto Suomi—La croisèe des destins* (1988), *source*: *Mikkola* (1997). *Reprinted with kind permission of Risto Suomi*. (d) *Wassily Kandinsky—Picture with a Black Arch* (1912), *source*: *WikiArt.org/public domain*. (d) *Nicolas Poussin—La Lamentation sur le Christ* (17*th century*), *source*: *The Yorck Project* (2002)*/public domain*. (e) *Pasi Tammi—Poem Forces to Kneel Down* (1999), *source*: *Anonymous* (2000). *Reprinted with kind permission of Pasi Tammi*.

FIG. 2. (*Left image*) *Koli landscape by Eero Järnefelt from the turn of* 19*th and* 20*th century.* (*Right image*) *Fixations* (*red spots*) *for Subject* 15, *a male non-novice. The duration of each fixation is proportional to the size of the spot.*

means, medians, and standard deviations of fixation durations for each subject. The number of fixations during the three minutes time period varies between 326 and 770, and the median fixation duration varies between 150 ms and 438 ms. The mean fixation duration is always greater than the median, indicating a right-skewed distribution of the fixation duration.

Short fixations are not included in our analysis because they are believed to be microsaccades, or eye movements within a fixation [Manor and Gordon (2003)]. The histogram on the left side of Figure 4 indicates that the duration distribution may be a mixture of very short durations and regular ones. When excluding the short fixations, the remaining durations seem to follow a gamma distribution. The

FIG. 3. *Areas of the Järnefelt painting that were visited most* (*left*) *and least* (*right*) *by the gaze of the subjects.* (*Left image*) *Top* 50% *white, top* 20% *yellow, top* 10% *orange, and top* 1% *red.* (*Right image*) *Bottom* 30% *blue and bottom* 10% *black.*

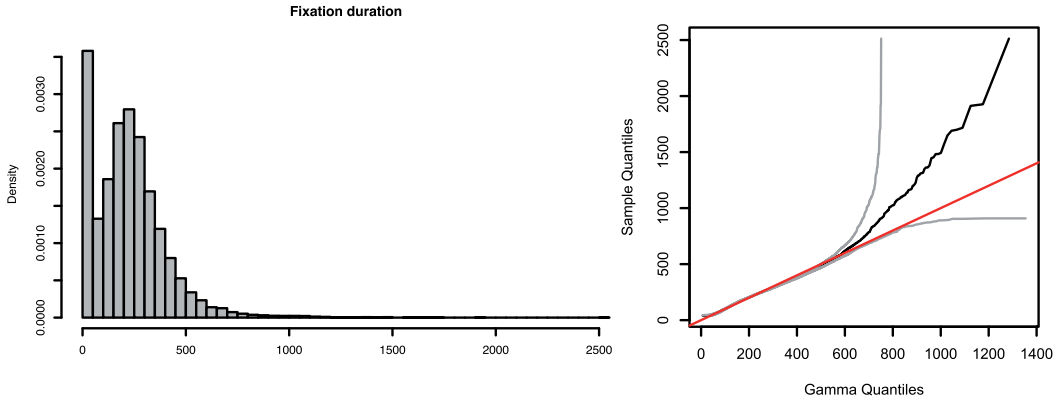

FIG. 4. (*Left panel*) *Distribution of fixation durations for all subjects when short fixations are included.* (*Right panel*) *Gamma quantile–quantile plot for fixation durations exceeding* 40 *ms with asymptotic* 95% *confidence region* (*grey*) *and the line* (*red*) *of unit slope through the origin corresponding to perfect agreement of empirical and theoretical quantiles.*

right panel shows a gamma quantile–quantile plot [Wilk and Gnanadesikan (1968)] of the regular (longer than 40 ms) fixation durations. The red line of slope 1 falls inside the simultaneous confidence band [Doksum and Sievers (1976)], indicating that the durations can be described well by a gamma distribution.

As mentioned earlier, the complete eye movement process consists of fixations and saccades, the latter being the rapid movements between the fixations. The supplemental article by Ylitalo, Särkkä and Guttorp (2016b) shows summary statistics for saccade durations and saccade lengths, that is, the distances between consecutive fixation locations. The distribution of saccade lengths as well as the distribution of saccade durations are skewed to the right, and are also well described by gamma distributions. The temporal dependence between fixation durations is very weak, as judged by the autocorrelation functions and ARIMA time series fitting.

2.3. *Temporal analysis.* Art is considered to be a subjective field. How one views artwork is individually unique, and reflects one's experience, knowledge, preference, and emotions. According to Locher's two-stage model [Locher (2006)], exploration of a picture starts with a global survey of the pictorial field in order to get an initial overall impression of the structural arrangement and semantic meaning of the composition. The second phase of an aesthetic episode consists of visual scrutiny or focal analysis of interesting pictorial features detected initially to satisfy cognitive curiosity and to develop aesthetic appreciation of the display. Some pioneering investigations into visual exploratory behavior of paintings by Buswell (1935) and Yarbus (1967), and subsequent studies on the informative details of an image by Antes (1974) and Mackworth and Morandi (1967), revealed that observers focus their gaze on specific areas of the image, rather than in a random fashion. The areas receiving high intensities of fixations were interpreted as

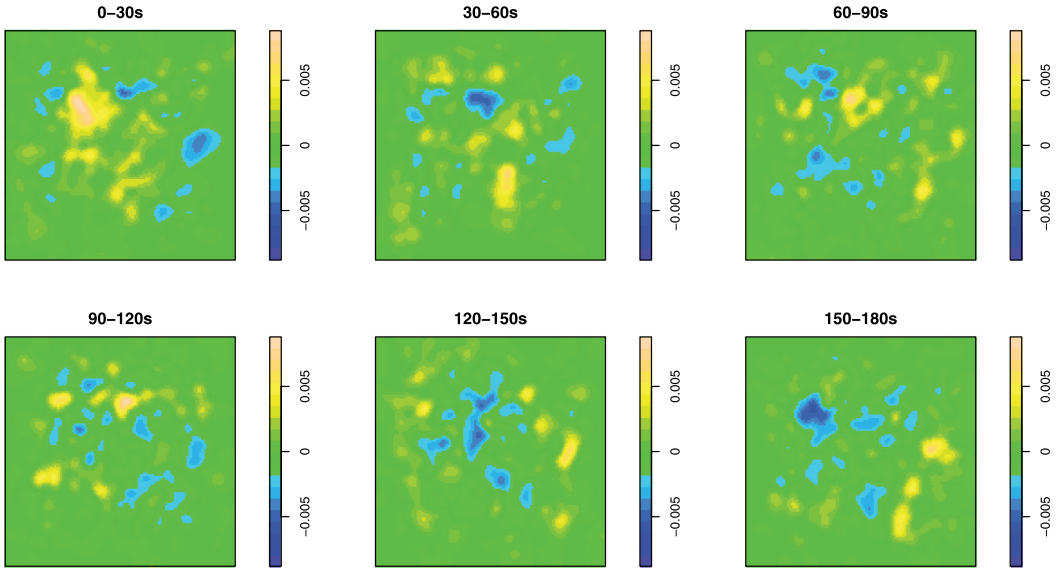

FIG. 5. *Residual intensity surfaces in* 30 *s intervals. The more yellow the color, the higher the intensity compared to the mean intensity surface; the more blue the color, the lower the intensity.*

guiding the observer's interest in informative elements of the image [Henderson and Hollingworth (1999)].

Since people do not look at all parts of the painting, but rather focus on some features of interest, the fixation process cannot be assumed to be spatially stationary. Furthermore, the features people look at may vary in time. To investigate the latter, we estimated the intensity surface of fixation locations based on the fixations of all subjects, and plotted residual intensities in 30 s intervals (Figure 5). The residual intensity is the difference between the estimated intensity surface and the mean intensity surface of the six intervals. The intensity in location $x \in I$, where I is the area of the painting, is estimated using the edge-corrected kernel estimator [see, e.g., Gatrell et al. (1996)]

$$\hat{\lambda}_h(x) = \frac{\sum_{i=1}^{n} h^{-2} K(h^{-1}(x - x_i))}{\int_I h^{-2} K(h^{-1}(x - u))\, du}, \tag{1}$$

where x_i, $i = 1, \ldots, n$, are the fixation locations, h a bandwidth (a smoothing parameter), and K a kernel function, here the standard two-dimensional Gaussian density. The bandwidth for the mean intensity estimation is selected by applying the cross-validation approach described in Diggle (1985) and Berman and Diggle (1989), and that bandwidth (here 17) is the same for all intervals.

One can notice in Figure 5 that during the first 30 s, the gaze is more concentrated on the tall pine tree on the left-hand side of the painting and after that more on the small tree in the middle than on average during the whole three minutes time period. During the second and the third minutes fixations are spread out a

little more than during the first minute. In addition, during the last minute the areas which were of interest at the beginning (the two trees) are now being avoided. Note that here we have only made some visual observations. For a more formal comparison of intensity surfaces we could use the test we introduce in Section 3 below.

We also investigate the difference in fixation duration distributions between the different 30 s time intervals. Figure 6 shows a graphical way to compare distributions, a shift function [Doksum and Sievers (1976)] defined as follows. If X has the cumulative distribution function F, written $X \sim F$, and Y the cumulative distribution function G, $Y \sim G$, then the shift function is defined as

$$\Delta(x) = G^{-1}\big(F(x)\big) - x,$$

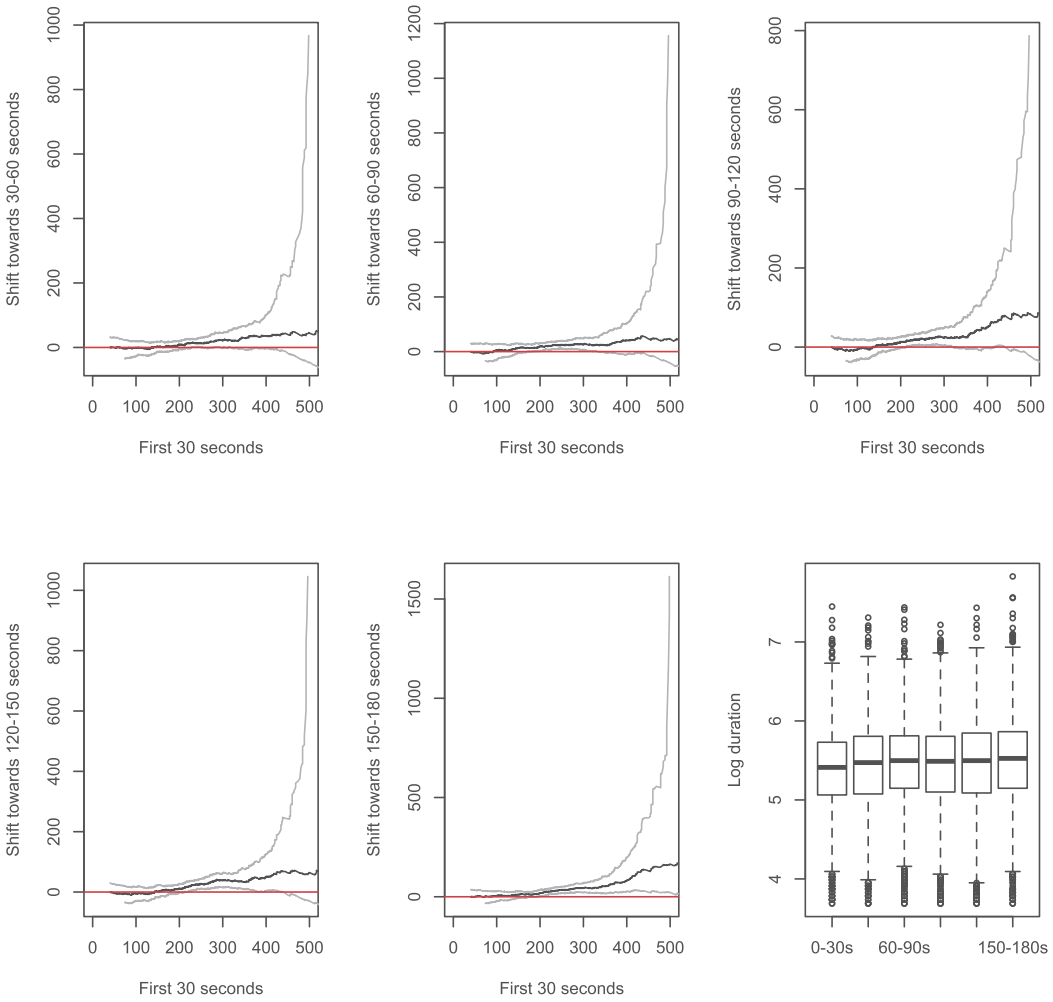

FIG. 6. *Fixation duration shift function estimates in* 30 *s intervals, relative to the first* 30 *s. The red line corresponds to equal distributions, and the grey lines form* 95% *asymptotic simultaneous confidence bands. The lower right panel is a box plot comparison of the distributions on a log scale.*

and has the property that $X + \Delta(X) \sim G$. The idea of the shift function is to compare two distributions based on how much one distribution function needs to be shifted in order to coincide with the other. In general, this may depend on the abscissa at which the comparison is made. We have the simplest case when the second distribution is simply the first shifted by a constant amount. The shift function is then just a horizontal line at the level of the shift (in particular, if the distributions are the same, then the shift function is a horizontal line at the level zero). If a location-scale model is appropriate, then the shift function is a straight line with slope related to the scale change. The shift function is easily estimated using empirical distribution functions, and simultaneous confidence intervals based on the distribution of the Kolmogorov–Smirnov statistic are given in Doksum and Sievers (1976). If a horizontal line at level 0 falls inside the band, then the two distributions are statistically indistinguishable. The shift function estimates in Figure 6 represent the comparison of the second through the sixth 30 s fixation distributions to the first 30 s distribution. It seems that the later time intervals tend to have slightly longer fixations between about 200 ms and 450 ms than the earlier intervals. This is also illustrated by the side-by-side box plots in the lower right-hand panel of the figure.

3. Comparison of novices and non-novices. In the previous section we introduced some tools to describe the fixation process. Now these tools are applied to compare the fixation processes of novices and non-novices. First, we compare the fixation intensity surfaces of the two groups visually, and then construct a test for more formal comparison. Second, we compare the fixation duration distributions in the two groups using shift plots. Finally, we investigate whether the fixation duration distributions change within the two groups if the painting is replaced by another.

3.1. *Visual comparison of intensity surfaces.* To compare the spatial patterns of fixations, we first estimate the overall intensity for each group by using equation (1); see the top row in Figure 7. We see that both groups concentrate on the middle part of the painting and do not look closely at the edges, and that the fixations of non-novices seem to be slightly more concentrated than the fixations of novices. Non-novices look a lot at the large pine and one of its branches (vertical alignment), while novices look at the same branch of the tree but continue to the right of the branch (horizontal alignment). We also plot the areas of the painting that are visited most (top 5% in yellow and top 1% in red) for each group (Figure 7, bottom row). This verifies the finding of non-novices' interest in the vertical trunk while novices have been concentrating more on the horizontal branch.

A more detailed investigation, where the time interval is divided into 30 s intervals [see supplemental article Ylitalo, Särkkä and Guttorp (2016c)], reveals that especially early on the non-novices are more concentrated on the large tree and its branch than the novices. The gaze of novices starts to spread out much earlier than

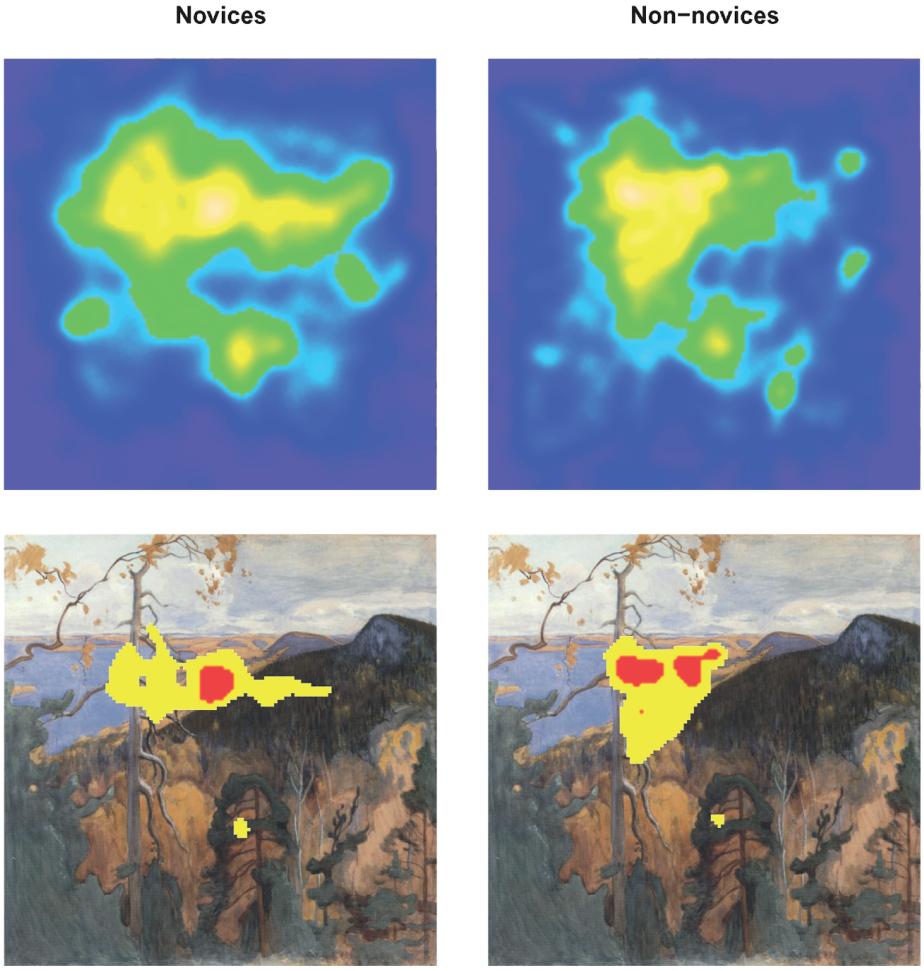

FIG. 7. *Top row*: *The estimated overall intensity surface with bandwidth* 20 *for novices* (*on the left*) *and with bandwidth* 16 *for non-novices* (*on the right*). *Bottom row*: *Top* 5% (*yellow*) *and top* 1% (*red*) *intensities*, *novices vs*. *non-novices*.

the gaze of non-novices and novices have more hot spots than non-novices. The gaze of non-novices mainly stays on the large pine tree and in the end also on the small tree on the right side of the large one, while novices investigate other areas of the painting as well.

3.2. *Test for comparing two intensity surfaces*. In order to test whether two intensity surfaces differ significantly from each other, that is, whether the subjects in the two groups have been looking at different areas of the painting, we apply the method developed by Kelsall and Diggle (1995a, 1995b). They considered a ratio of two kernel density estimates and used a Monte Carlo test for testing if the ratio is constant, that is, whether the kernel estimated intensity surfaces can be considered similar. This ratio is used in epidemiology to explore whether some diseased cases are randomly located among healthy controls and is therefore often

called a relative risk function. The logarithmic relative risk function is defined as $\rho(x) = \log(\lambda_1(x)/\lambda_2(x))$, where λ_i's are computed using equation (1), and can reveal areas where the intensity of fixations differs between novices and non-novices.

By conditioning on the observed number of fixations, we may view the fixation locations of novices and non-novices as independent samples from the densities $f_1(x)$ and $f_2(x)$, where

$$\hat{f}_i(x) = \frac{\hat{\lambda}_i(x)}{\int_I \hat{\lambda}_i(x)\,dx},$$

for $i = 1, 2$. We can now define a logarithmic density ratio

$$r(x) = \log \frac{f_1(x)}{f_2(x)} = \rho(x) - c,$$

where $c = \log(\frac{\int_I \lambda_1(x)\,dx}{\int_I \lambda_2(x)\,dx})$ [see, e.g., Wakefield, Kelsall and Morris (2000), Kelsall and Diggle (1995b)]. We have

$$\hat{r}(x) = \log \frac{\hat{f}_1(x)}{\hat{f}_2(x)},$$

which contains all the information about the spatial variation in $\rho(x)$.

According to Kelsall and Diggle (1995b), the choice of the kernel function is not critical, but choosing the smoothing parameter h is. Furthermore, it is not obvious whether one should use the same bandwidth for both densities when computing the ratio. Kelsall and Diggle (1995b) suggest that if the densities are expected to be nearly equal, then the bandwidths should be equal as well to reduce bias. They also warn that if the sample sizes are very different, then one can get poor results when using the same bandwidth for both samples. Furthermore, Bailey and Gatrell (1995) suggest that the kernel estimate in the denominator should be deliberately oversmoothed (a larger bandwidth should be used). In addition, Hazelton (2007) shows that edge correction terms are not needed if a common bandwidth is used.

We test the null hypothesis that the intensity surfaces are equal, that is, that the logarithmic density ratio $r(x) = 0$, by the following Monte Carlo test [Barnard (1963)]. Ten subjects were randomly selected from the twenty subjects to form the novice group and the remaining ten subjects formed the non-novice group. (Hence, the possible dependence between the fixation locations within a subject was taken into account.) By doing this random grouping m times, we obtain m log-density surfaces $\hat{r}_1, \ldots, \hat{r}_m$ under the null hypothesis. As introduced by Kelsall and Diggle (1995b), we use the test statistic

$$T_j = \int_I \hat{r}_j(x)^2\,dx, \qquad j = 0, \ldots, m.$$

The p-value is then $p = \frac{k+1}{m+1}$, where k is the number of test statistics T_j which are larger than or equal to the observed value of the test statistic, T_0.

Since the number of fixations vary quite a lot between the subjects, we estimate the bandwidths separately for the two groups by using the cross-validation method by Diggle (1985) and Berman and Diggle (1989) with Ripley's isotropic edge correction [Ripley (1988)]. The estimated bandwidth for the novices was approximately 20 and for the non-novices 16. We obtained $T_0 = 59{,}604$ and $p = 0.108$ after random grouping of subjects 10,000 times. This means that the difference between the intensity surfaces of novices and non-novices is not statistically significant even though there are some dissimilarities in the intensities; see Figure 7. The largest differences between the intensities of the two groups are near the edges of the painting, where there are very few fixations of either group.

We made the same comparisons for the other five paintings and obtained similar results; see supplemental article Ylitalo, Särkkä and Guttorp (2016d). None of the tests gave any significant result, and Fisher's combined probability test resulted in $p = 0.392$ ($\chi^2 = 12.685$, $df = 12$). Therefore, we did not find any overall differences when comparing the intensities of the two groups. However, it seems that the more abstract paintings could reveal some differences between novices and non-novices; see results related to Kandinsky's painting in Ylitalo, Särkkä and Guttorp (2016d).

3.3. *Comparing duration distributions.* To compare the fixation duration distributions of the novice and non-novice groups, we again look at the shift plot; for the Järnefelt painting see Figure 8. Here, the novices have fewer short fixations (duration less than 125 ms), and more fixations between about 175 and 600 ms than non-novices. Similar results were found for the remaining five paintings; see Ylitalo, Särkkä and Guttorp (2016d). This is in line with the early findings of Buswell (1935), and may mean that novices more rarely only glance at a point on the painting but rather spend more time on each point compared to non-novices.

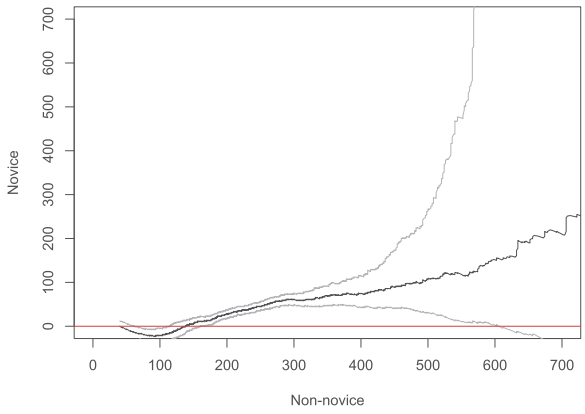

FIG. 8. *Shift plot of non-novice distribution vs. novice distribution for fixation duration. The red line indicates equal distributions, and the grey lines form simultaneous asymptotic 95% confidence bands.*

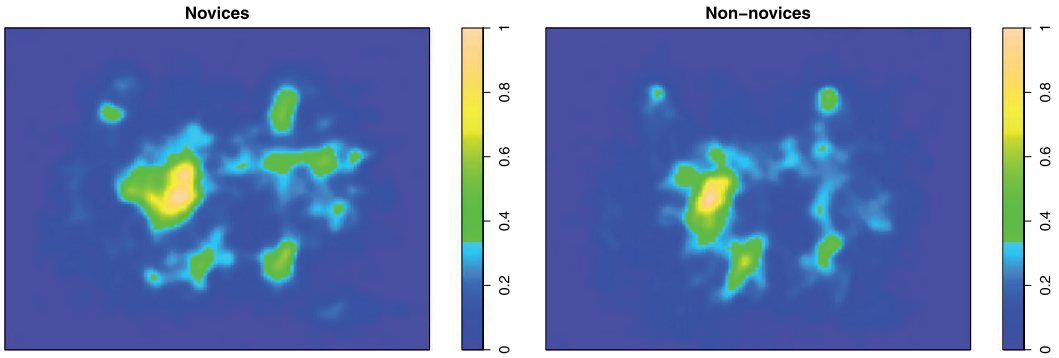

FIG. 9. *The estimated overall intensity surface with fixed bandwidth* 13 *for novices* (*on the left*) *and for non-novices* (*on the right*) *for the Monet painting.*

3.4. *Pairwise comparison of paintings.* Since several paintings are available, we are also able to investigate whether the fixation process is similar for different paintings, that is, whether the fixation process stays stable when looking at paintings. We are particularly interested in whether we detect any differences between the behavior in the novice and non-novice groups. We started by comparing the Järnefelt painting and the painting Terrace at Sainte-Adresse by Claude Monet (1867) shown in Figure 1(b) since both are landscapes, and hence similar to each other. The estimated intensity surface of the fixation locations for Monet's painting is also shown separately for novices and non-novices in Figure 9. The intensity surfaces of the two groups look fairly similar; see also the formal test in Ylitalo, Särkkä and Guttorp (2016d).

Since the pictures are completely different, the intensity surfaces of fixations for the paintings are also different and comparing them would be meaningless. However, it is of interest to compare the fixation duration distributions of novices and non-novices between two paintings in order to see whether the way of looking at paintings remains stable between different paintings. The shift function comparisons of the duration distributions for novices and non-novices are plotted in Figure 10 for the Järnefelt–Monet pair. We can see that for non-novices there is no significant difference between the paintings, while for novices there are more fixations that last 200–350 ms on the Järnefelt painting than on the Monet painting. We made similar comparisons for the other pairs of paintings as well; see supplemental article Ylitalo, Särkkä and Guttorp (2016e). Here, only the results based on the comparisons of the fixation duration distributions for the most similar pairs of paintings, namely, Suomi and Kandinsky which are abstract paintings [see Figures 1(c) and (d)], and Poussin and Tammi which have people in focus [see Figures 1(e) and (f)], are included. For novices, we see a difference for each three pairs of paintings. For non-novices, however, we can only see a clear difference for the Suomi–Kandinsky pair, but there is no difference for the other two pairs of paintings. This indicates that the non-novices, being used to looking at paintings,

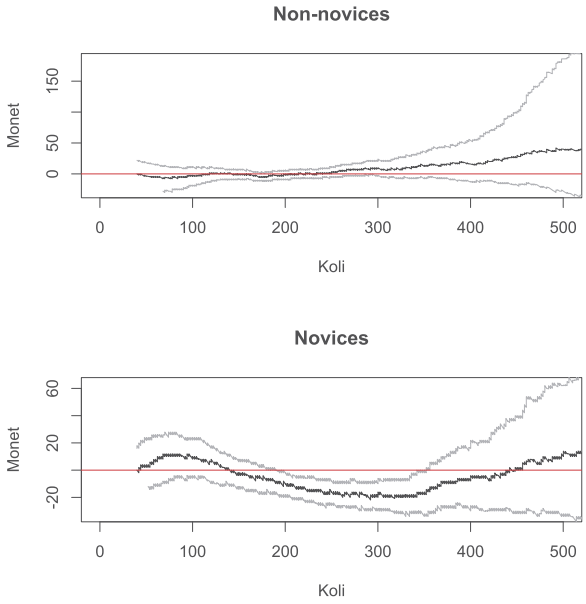

FIG. 10. *Shift function comparison of fixation duration distributions between the Järnefelt and Monet paintings for non-novices* (*upper panel*) *and novices* (*lower panel*). *The red lines correspond to equal distributions and the grey lines form* 95% *asymptotic simultaneous confidence bands*.

are more consistent in their fixation durations between paintings, but for novices, the duration of fixations depends on the painting.

4. Stochastic model. Above, we used intensity surfaces and shift plots to compare the eye movements of novices and non-novices. Next, we will construct a simple spatio-temporal model that describes the dynamics of the fixation process. We will fit the model separately to the novice and non-novice groups for the Järnefelt painting and see if we can find some further differences between the groups by comparing the fitted models in terms of functional summary statistics and their model-based simultaneous envelopes.

To build a spatio-temporal model for the fixation process, we need to model fixation locations, fixation durations, saccade durations, and saccade lengths. (Saccade durations are needed to fill the whole three minutes time period.) As building blocks we use the tools discussed in Section 2. Fixation locations are modeled as a realization of a spatial point process, which is characterized by its intensity function. Furthermore, appropriate distributions can be fitted to the fixation and saccade durations and for saccade lengths. After having estimated the intensity surface and the distributions mentioned above, we can simulate from the resulting model and compare the behavior of the model to the behavior in the data, and finally compare the models fitted to the two groups.

A natural reference or null model for spatial point pattern data, locations of fixations in our case, is the homogeneous Poisson process, where the points are

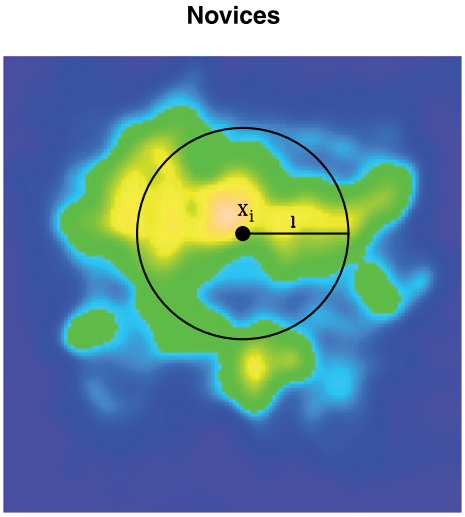

FIG. 11. *The next fixation location is chosen from the black circle at distance l from the current fixation x_i according to the estimated intensity surface.*

located uniformly and independently of each other on the study area. In our situation, however, only the more interesting parts of the painting are visited, and we cannot assume that the final set of locations of fixations is a realization of a homogeneous process. A homogeneous Poisson process is therefore not a realistic reference process for the fixation locations. Next, we define our own reference model for the fixation process.

Given an intensity surface, the first fixation location is drawn from the probability distribution on the observation window by normalizing the intensity to a bivariate distribution. The length l of the next saccade is drawn from a saccade length distribution for the group in question. Then, the exact location of the new fixation is decided by taking a point at distance l away from the current fixation location according to the intensity surface. In other words, the new location is chosen from the conditional intensity of points at the given distance from the current location; see Figure 11.

Concerning the temporal evolution, it seems natural to draw each fixation duration independently from the distribution of durations.

5. Model fit and comparison between novices and non-novices for the Järnefelt painting. In what follows, we will first give the particular choices of the distributions mentioned in the previous section, and introduce some summary statistics that could be used to assess the goodness of fit of the fitted models. Finally, we fit the model to the data for the Järnefelt painting, separately for novices and non-novices, simulate 200 realizations of each process, and check the goodness of fit of the fitted models by comparing the summary statistics estimated from the data to those estimated from the simulations.

5.1. *Choice of distributions*. The intensity surfaces of the fixation locations are estimated separately for each group and the bandwidth is selected by applying the cross-validation approach implemented in the R package `spatstat` [Baddeley, Rubak and Turner (2015)]. The same bandwidth is used for the data and for the simulations. The first location is drawn from the intensity surface estimated from the locations of the first fixations of the subjects in the group. Alternatively, the intensity surface estimated based on all the fixation locations in the group could be used. We believe that the choice of the first fixation location is not crucial.

Given the summary statistics computed from the data (Section 2.2), an appropriate distribution for the duration of fixations and the duration of saccades is the gamma distribution. Therefore, we model the fixation durations by sampling from the estimated gamma distribution of fixation durations in the group. The duration of each saccade is modeled by drawing from the common distribution of saccade durations estimated from all the subjects, as saccades are involuntary, once started, and therefore not expected to vary depending on viewing experience. The intensity surface, as well as the fixation duration and saccade duration distributions, stay the same throughout the observation period.

For saccade lengths we first fit a Gamma distribution separately for novices and non-novices and then use truncated versions of these distributions in order to avoid jumps outside the painting. The truncation point depends on the current location: if the process is at location x_i, then the longest jump it can take is to the furthest corner of the painting. This distance is called $l_{i,\max}$ and serves as the truncation point for the truncated gamma distribution from which the next jump is sampled. We noticed, however, that the truncated gamma distribution alone does not catch the long jumps that the gaze makes when the person moves her/his attention to another location. Therefore, we include such long jumps into the model and define the distribution for saccade lengths to be a mixture of two distributions: When the process is at location x_i, the length of the next jump is sampled from the uniform distribution $U(l_{i,\max}/2, l_{i,\max})$ with probability p and from the truncated Gamma distribution with truncation point $l_{i,\max}$ with probability $1 - p$. After some experimentation, we fixed the probability p for sampling the saccade length from the uniform distribution above to 0.2 for both groups.

5.2. *Summary statistics*. To get some idea of how the gaze jumps between different areas of the painting in the data and in the simulations, we divide the painting into quarters and compute the transition frequencies between them. Each quarter is called a state and identified with numbers 1–4 (from upper left to lower right quarter).

In order to see how much of the painting is viewed, we use two functional summary statistics, namely, convex hull coverage and ball union coverage defined as follows. Let, as before, I represent the observation window and $x_i \in I$ be the

ith fixation of the fixation process $\{X(t)\}$. An ordered set of n_t fixations up to a fixed time t is denoted by $x = \{x_1, \ldots, x_{n_t}\}$. The convex hull of this point set is

$$C_x(t) = \left\{ \sum_{i=1}^{n_t} \alpha_i x_i : \alpha_i \geq 0 \text{ for all } i \text{ and } \sum_{i=1}^{n_t} \alpha_i = 1 \right\}$$

and the relative area of the convex hull is denoted by

$$AC_x(t) = \begin{cases} 0, & \text{if } n_t < 3, \\ \dfrac{\text{area}(C_x(t))}{\text{area}(I)}, & \text{if } n_t \geq 3. \end{cases}$$

This is called the convex hull coverage of the fixation process. Hence, the convex hull coverage is computed every time a new fixation appears.

Let $b(x_i, R)$ be a ball of fixed radius R centered in a fixation $x_i \in I$. Define

$$U_x(t) = \left\{ \bigcup_{i=1}^{n_t} b(x_i, R) \cap W \right\},$$

and call

$$AU_x(t) = \frac{\text{area}(U_x(t))}{\text{area}(I)}$$

the ball union coverage of the fixation process.

In addition, to measure the total length of the distance the gaze has moved, we use the scanpath length function [Noton and Stark (1971)], which is the sum of the saccade lengths up to time t. Hence, the scanpath length can be defined as

$$L(t) = \sum_{i=1}^{n_t - 1} l_i \mathbf{1}(t_{i+1} \leq t),$$

where t_{i+1} is the time when the fixation $i + 1$ at location x_{i+1} takes place and l_i is the length of the jump just before that fixation. That length is the Euclidean distance between two successive fixations, that is, the length of the saccade.

5.3. *Model fit.* The estimated model-based intensity surfaces for the groups are different; see Figure 12. As mentioned earlier, both groups look a lot at the large pine tree and the top of the small tree next to it. The novice group tends to follow the branch of the large tree, while the non-novice group follows the trunk.

The convex hull coverage, ball union coverage with disc radius 35 pixels, and scanpath length for non-novices and novices are shown and compared to the models in Figure 13. The 95% simultaneous (rank) envelopes were created by using the R package `spptest` [Myllymäki et al. (2015, 2016)]. Figure 13 shows that the coverage statistics based on the model describe the data quite well. The model-based scanpath length is, however, less variable and somewhat shorter in the simulations than in the data. It seems that the model is not able to catch all variation

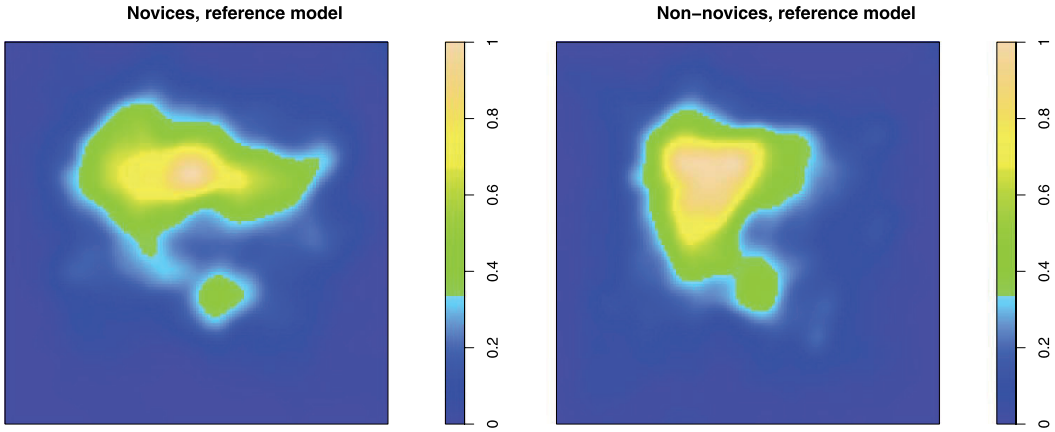

FIG. 12. *Scaled overall intensity surfaces of the reference model for novices* (*left*) *and for non-novices* (*right*).

related to the scanpath length. In the beginning (during the first 30 seconds for novices and during the first minute for non-novices), the model describes the data well, but later on variation in the data becomes too large to be caught by the model.

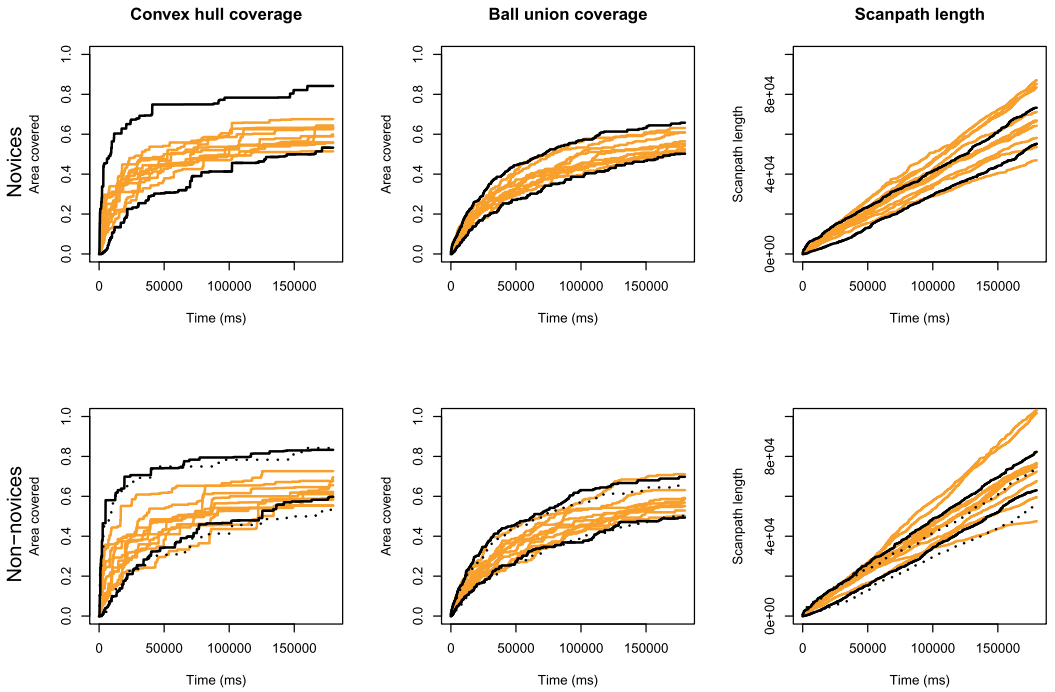

FIG. 13. *Convex hull coverage* (*left*), *ball union coverage* (*middle*), *and scanpath length* (*right*) *curves for novices* (*top*) *and non-novices* (*bottom*). *Orange lines represent the subjects and black solid lines represent* 95% *simultaneous envelopes estimated from* 200 *simulated realizations of the reference model of the group in question. Dotted lines* (*bottom*) *represent* 95% *simultaneous envelopes for novices.*

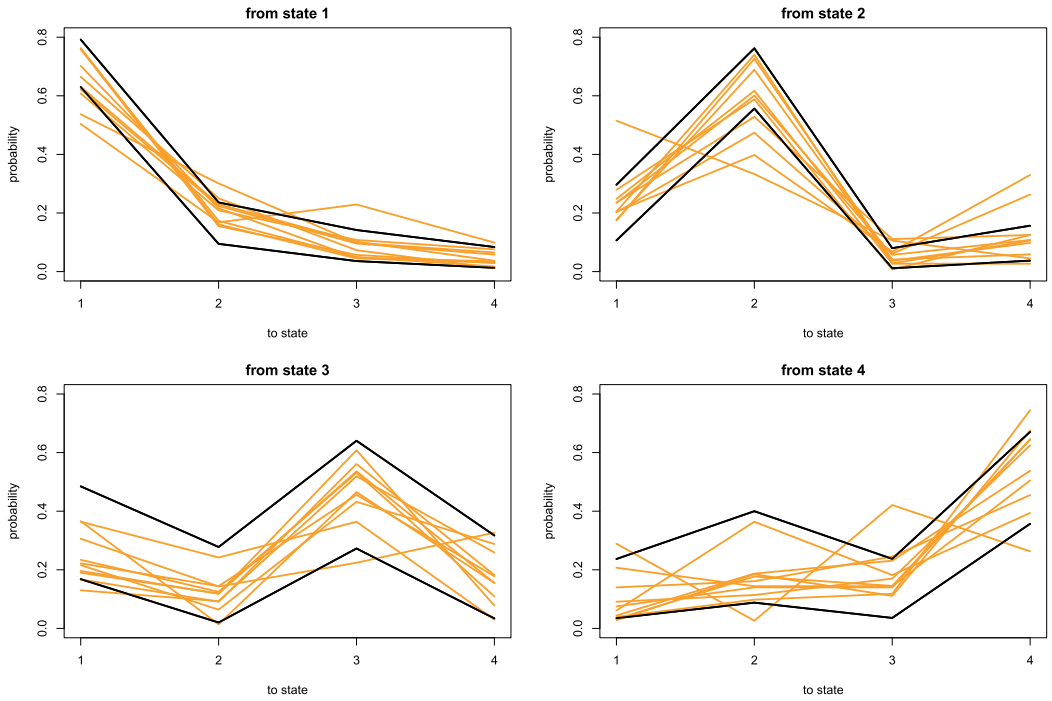

FIG. 14. *Transition probabilities for novices with* 95% *simultaneous envelopes estimated from* 200 *simulated realizations of the reference model.*

One should also note that the scanpath length is strongly affected by the number of fixations and saccade lengths, which both vary quite a lot in our data; see supplemental articles Ylitalo, Särkkä and Guttorp (2016a) and Ylitalo, Särkkä and Guttorp (2016b).

For both the novice data and the non-novice data, the estimated transition probabilities between different quarters of the painting indicate that the model does an adequate job (Figures 14 and 15). Recall that the quarters of the painting are called states and numbered from 1 to 4 (from upper left to lower right quarter).

The model seems to fit quite well to the fixation process for both groups. The fitted models are quite similar, but also some differences can be found. As can be seen in Figure 13, the model-based coverage summaries are quite similar for the two groups. However, the modeled gaze of non-novices seems to move slightly more than the gaze of novices according to the scanpath length (see Figure 13, bottom right plot). This is somewhat surprising, since the intensity surface of non-novices seems to be more compact than the one of novices, which would indicate that the gaze of non-novices takes shorter jumps in smaller areas than the gaze of novices. However, fixation durations of novices are longer than the ones of non-novices in general, which means that non-novices tend to make more jumps during the three minute period compared to novices. Thus, it is possible that the gaze of non-novices makes a larger number of short jumps than the gaze of novices, which results in the gaze traveling a longer path than the gaze of novices.

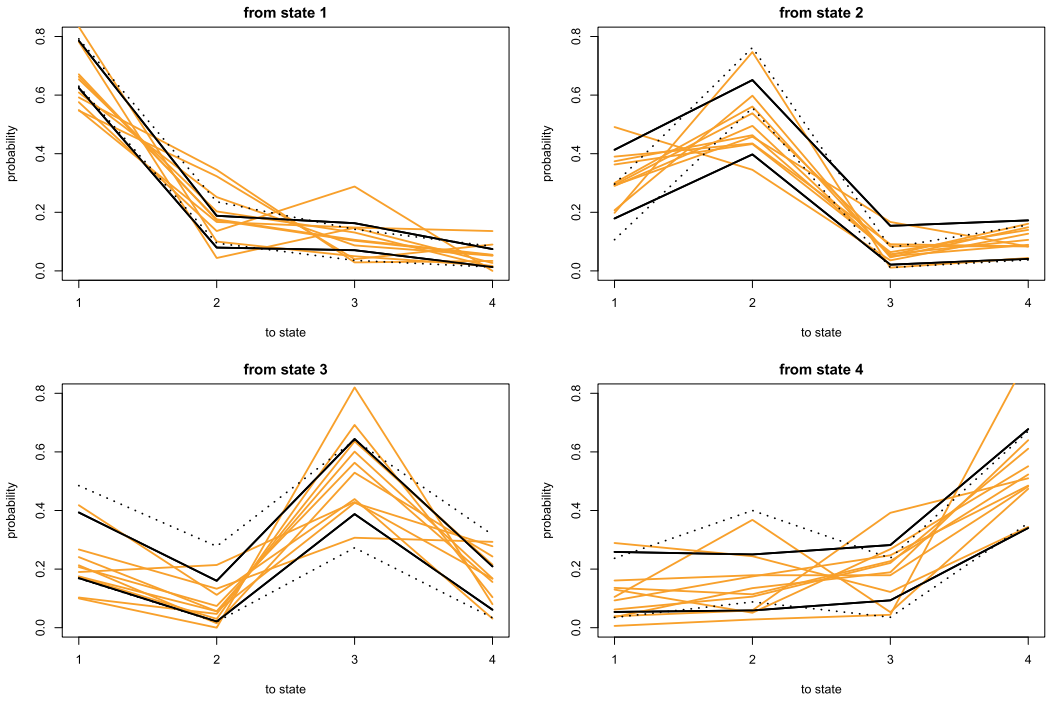

FIG. 15. *Transition probabilities for non-novices with* 95% *simultaneous envelopes estimated from* 200 *simulated realizations of the reference model. Dotted lines represent* 95% *simultaneous envelopes for novices.*

6. Conclusions and discussion. In this study we have analyzed eye movement data from twenty subjects who had looked at six different paintings. We have been particularly interested in whether non-novices and novices look at paintings in the same way, both when it comes to which parts of the painting they look at and how long the gaze typically stays in one spot. By regarding the fixation process as a spatio-temporal point process and by using a new set of tools we can study the differences from a slightly different perspective than what has been done earlier. Our main purpose has been to see how useful our new approach is and, therefore, we have mainly concentrated on one of the six paintings that we have had available, namely, the Järnefelt painting *Koli landscape*. However, some results based on the other paintings are included.

By looking at the intensity surfaces of the point patterns formed by the locations of fixations we have been able to see some differences between the eye movements of novices and non-novices on the Koli painting. The area that is visited most is more concentrated and more stable in time for the non-novice group than for the novice group. We also constructed a test to compare the intensity surfaces of novices and non-novices. Even though the observed difference was not statistically significant, there was some indication that the intensities cannot be considered equal. By comparing fixation duration distributions instead of bare means we were able to see a clear difference between the two groups under study. Non-novices

tend to have more short fixation durations than novices, meaning that they may only glance at some areas of the painting which novices do not tend to do. This was confirmed by the analysis based on the other paintings.

To have some idea whether the painting itself affects how novices and non-novices look at it, we compared the fixation processes of each group on three pairs of paintings, Järnefelt–Monet, Suomi–Kandinsky, and Poussin–Tammi. The intensity surface of fixations vary of course from painting to painting and it is not meaningful to compare them on different paintings, but it does make sense to compare the duration distributions. An interesting observation was that in the non-novice group the distribution of duration of fixations is more consistent from painting to painting than in the novice group.

Ideally, to be able to have a more detailed comparison of the eye movements of novices and non-novices, we would like to be able to describe the complete fixation process on a painting, that is, where, when, and how long the gaze of a person stays in different parts of the painting. Therefore, we have introduced a simple reference model for the fixation process. The model is a spatio-temporal point process model, where we used the fixation intensity and duration distributions estimated from the data. In order to restrict how far the gaze usually jumps, we also estimated the saccade length distribution from the data. To mimic the tendency that after having stayed on one area of the target painting for some time the gaze jumps to another area on the painting, the model suggests a long jump with some (small) probability p.

We fitted the model to the non-novice and novice groups separately for the eye movements on the Järnefelt painting and saw that the model fits quite well for both groups, except when measured by the scanpath length. The structure of the fixation process (described by our model) is similar in nature for non-novices and novices. The model-based simultaneous envelopes for the functional summary statistics can be used like confidence intervals to compare the two groups. However, since the envelopes overlap, we were not able to reveal any significant differences between the groups based on the model-based summaries.

Our idea here was to find a rather simple model that captures the most important features of the fixation process, and the reference model is quite good for this purpose. However, if the goal is to understand the complete dynamics of the fixation process, then the model constructed here is not good enough since it does not capture that the fixation process is changing in time. We noticed, for example, in Section 2 that both the intensity surface and the fixation duration distribution vary in time. Despite this, we used the same intensity surface and the same gamma distributions for the fixation durations and saccade lengths during the whole time period. We made some experiments by using six intensity surfaces (for each 30 second interval) and two distributions for saccade lengths (one for the first 30 seconds and one for the rest of the three minute time interval) in the reference model, but these modifications did not improve the goodness of fit of the model.

One possible way to improve the model could be to take the order of fixations into account. The fixation process could be regarded as a realization of a sequential point process which can be used to access the nonstationarity of the fixation process. The first author of this paper is currently working on that issue [see Penttinen and Ylitalo (in press)].

To conclude, our new set of tools allowed us to make more detailed comparisons between how novices and non-novices look at paintings than reported in the literature. As far as we know, such comparisons of fixation intensity surfaces, duration distributions, and coverages have not been presented earlier in eye movement studies.

Acknowledgments. The authors are grateful to Pertti Saariluoma, Sari Kuuva, María Álvarez Gil, Jarkko Hautala, and Tuomo Kujala for providing the data and to Antti Penttinen for helpful comments and discussions. The authors thank the two anonymous reviewers, Associate Editor, and Editor for their valuable comments that helped to significantly improve the paper.

The mobility funding provided by the University of Jyväskylä for the first author (AKY) is highly appreciated.

SUPPLEMENTARY MATERIAL

Supplement A: Statistics for fixations inside the picture (DOI: 10.1214/16-AOAS921SUPPA; .pdf). Table includes information about the fixations, such as total number of fixations and mean fixation duration, for each subject.

Supplement B: Statistics for saccades inside the picture (DOI: 10.1214/16-AOAS921SUPPB; .pdf). Table includes information about the saccades, such as mean saccade duration and mean saccade length, for each subject.

Supplement C: The most visited areas during the six 30 second intervals for novices and non-novices (DOI: 10.1214/16-AOAS921SUPPC; .pdf). Figures show the top 5% and top 1% intensities in 30 second intervals (0–30 s), (30–60 s), (60–90 s), (90–120 s), (120–150 s), and (150–180 s) for novices and non-novices.

Supplement D: Results for the comparison of novices and non-novices for all six paintings (DOI: 10.1214/16-AOAS921SUPPD; .pdf). Results for the intensity surface and fixation duration distribution comparisons between novices and non-novices for all six paintings used in the experiment.

Supplement E: Results for the groupwise fixation duration distribution comparisons for three pairs of paintings (DOI: 10.1214/16-AOAS921SUPPE; .pdf). Results for the fixation duration distribution comparisons within novice and non-novice groups for the three pairs of paintings: Järnefelt–Monet, Kandinsky–Suomi, and Poussin–Tammi.

REFERENCES

ANONYMOUS (2000). Brochure of Art Centre Salmela. Mäntyharju, Finland.

ANTES, J. R. (1974). The time course of picture viewing. *J. Exp. Psychol.* **103** 62–70.

BADDELEY, A., RUBAK, E. and TURNER, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC Press, London. Available at http://www.taylorandfrancis.com/books/details/9781482210200/.

BAILEY, T. C. and GATRELL, A. C. (1995). *Interactive Spatial Data Analysis*. Longman, Harlow.

BARLOW, H. B. (1952). Eye movements during fixation. *J. Physiol.* **116** 290–306.

BARNARD, G. A. (1963). Contribution to the discussion of professor Bartlett's paper. *J. Roy. Statist. Soc. Ser. A* **25** 294.

BARTHELMÉ, S., TRUKENBROD, H., ENGBERT, R. and WICHMANN, F. (2013). Modelling fixation locations using spatial point processes. *J. Vis.* **13** 1–34.

BERMAN, M. and DIGGLE, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *J. Roy. Statist. Soc. Ser. B* **51** 81–92. MR0984995

BUSWELL, G. T. (1935). *How People Look at Pictures—A Study of the Psychology of Perception in Art*. The Univ. Chicago Press, Chicago, IL.

DIGGLE, P. J. (1985). A kernel method for smoothing point process data. *J. R. Stat. Soc. Ser. C.* **34** 138–147.

DOKSUM, K. A. and SIEVERS, G. L. (1976). Plotting with confidence: Graphical comparisons of two populations. *Biometrika* **63** 421–434. MR0443210

DUCHOWSKI, A. T. (2002). A breadth-first survey of eye tracking applications. *Behav. Res. Methods Instrum. Comput.* **34** 455–470.

ENGBERT, R., TRUKENBROD, H. A., BARTHELMÉ, S. and WICHMANN, F. A. (2015). Spatial statistics and attentional dynamics in scene viewing. *J. Vis.* **15** 1–17.

FINDLAY, J. M. (2009). Saccadic eye movement programming: Sensory and attentional factors. *Psychol. Res.* **73** 127–135.

GATRELL, A. C., BAILEY, T. C., DIGGLE, P. J. and ROWLINGSON, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers, New Series* **21** 256–274.

HAZELTON, M. L. (2007). Kernel estimation of risk surfaces without the need for edge correction. *Stat. Med.* **27** 2269–2272.

HENDERSON, J. M. and HOLLINGWORTH, A. (1999). High-level scene perception. *Annu. Rev. Psychol.* **50** 243–271.

KELSALL, J. E. and DIGGLE, P. J. (1995a). Kernel estimation of relative risk. *Bernoulli* **1** 3–16. MR1354453

KELSALL, J. E. and DIGGLE, P. J. (1995b). Non-parametric estimation of spatial variation in relative risk. *Stat. Med.* **14** 2335–2342.

KINSLER, V. and CARPENTER, R. H. S. (1995). Saccadic eye movements while reading music. *Vis. Res.* **35** 1447–1458.

KOMOGARTSEV, O. V., RYU, Y. S. and KOH, D. H. (2009). Quick models for saccade amplitude prediction. *Journal of Eye Movement Research* **1** 1–13.

KRISTJANSON, A. F. and ANTES, J. R. (1989). Eye movement analysis of artists and nonartists viewing paintings. *Vis. Arts Res.* **15** 21–30.

LOCHER, P. J. (2006). The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. *Psychol. Sci.* **48** 106–114.

MACKWORTH, N. H. and MORANDI, A. J. (1967). The gaze selects informative details within pictures. *Atten. Percept. Psychophys.* **2** 547–552.

MANOR, B. R. and GORDON, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *J. Neurosci. Methods* **128** 85–93.

MIKKOLA, K., ED. (1997). Risto Suomi. Amos Anderson Art Museum, publications, new series, no 25.

MYLLYMÄKI, M., MRKVICKA, T., GRABARNIK, P., SEIJO, H. and HAHN, U. (2016). Global envelope tests for spatial processes. *J. Roy. Statist. Soc. Ser. B*. DOI:10.1111/rssb.12172.

MYLLYMÄKI, M., GRABARNIK, P., SEIJO, H. and STOYAN, D. (2015). Deviation test construction and power comparison for marked spatial point patterns. *Spat. Stat.* **11** 19–34. MR3311854

NAGASAWA, S., YIM, S. and HONGO, H. (2005). Feasibility study on marketing research using eye movement: An investigation of image presentation using an "eye camera" and data processing. *J. Adv. Comput. Intell. Intell. Informa.* **9** 440–452.

NOTON, D. and STARK, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vis. Res.* **11** 929–942.

PENTTINEN, A. and YLITALO, A.-K. (in press). Deducing self-interaction in eye movement data using sequential spatial point processes. *Spatial Statistics*. DOI:10.1016/j.spasta.2016.03.005.

THE YORCK PROJECT (2002). 10.000 Meisterwerke der Malerei, DVD-ROM, 2002. ISBN 3936122202. Distributed by DIRECTMEDIA Publishing GmbH.

RAYNER, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **124** 372–422.

RAYNER, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol., A Hum. Exp. Psychol.* **62** 1457–1506.

RIPLEY, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge Univ. Press, Cambridge. MR0971986

SUNDELL, D. (1986). *Eero Järnefelt (1863–1937), Retretti 25.5.–21.9.1986*. Retretti.

VOGT, S. and MAGNUSSEN, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception* **36** 91–100.

WADE, N. J. (2010). Pioneers of eye movement research. *I-Perception* **1** 33–68.

WAKEFIELD, J. C., KELSALL, J. E. and MORRIS, S. E. (2000). Clustering, cluster detection, and spatial variation in risk. In *Spatial Epidemiology: Methods and Applications* (P. Elliot, J. C. Wakefield, N. G. Best and D. J. Briggs, eds.) 128–152. Oxford Univ. Press, London.

WILK, M. B. and GNANADESIKAN, R. (1968). Probability plotting methods for the analysis of data. *Biometrika* **55** 1–17.

YARBUS, A. L. (1967). *Eye Movements and Vision*. Plenum Press, New York.

YLITALO, A.-K., SÄRKKÄ, A. and GUTTORP, P. (2016a). Supplement to "What we look at in paintings: A comparison between experienced and inexperienced art viewers." DOI:10.1214/16-AOAS921SUPPA.

YLITALO, A.-K., SÄRKKÄ, A. and GUTTORP, P. (2016b). Supplement to "What we look at in paintings: A comparison between experienced and inexperienced art viewers." DOI:10.1214/16-AOAS921SUPPB.

YLITALO, A.-K., SÄRKKÄ, A. and GUTTORP, P. (2016c). Supplement to "What we look at in paintings: A comparison between experienced and inexperienced art viewers." DOI:10.1214/16-AOAS921SUPPC.

YLITALO, A.-K., SÄRKKÄ, A. and GUTTORP, P. (2016d). Supplement to "What we look at in paintings: A comparison between experienced and inexperienced art viewers." DOI:10.1214/16-AOAS921SUPPD.

YLITALO, A.-K., SÄRKKÄ, A. and GUTTORP, P. (2016e). Supplement to "What we look at in paintings: A comparison between experienced and inexperienced art viewers." DOI:10.1214/16-AOAS921SUPPE.

A.-K. YLITALO
DEPARTMENT OF MUSIC
AND
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF JYVASKYLA
P.O. BOX 35
FI-40014
FINLAND
E-MAIL: anna-kaisa.ylitalo@jyu.fi

A. SÄRKKÄ
MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
AND UNIVERSITY OF GOTHENBURG
SE-412 96 GOTHENBURG
SWEDEN
E-MAIL: aila@chalmers.se

P. GUTTORP
DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WASHINGTON 98195-4322
USA
AND
NORWEGIAN COMPUTING CENTER
P.O. BOX 114 BLINDERN
NO-0314 OSLO
NORWAY
E-MAIL: peter@stat.washington.edu

SUPPLEMENT A TO "WHAT WE LOOK AT IN PAINTINGS: A COMPARISON BETWEEN EXPERIENCED AND INEXPERIENCED ART VIEWERS"

By Anna-Kaisa Ylitalo*, Aila Särkkä† and Peter Guttorp‡

University of Jyvaskyla, Chalmers University of Technology and University of Gothenburg†, University of Washington and Norwegian Computing Center* ‡

Statistics for fixations inside the picture.

Subject	total number of fix. inside the picture	number of short fix. inside the picture	mean fixation duration (ms)	median fixation duration (ms)	sd fixation duration (ms)
1 NN, F	653	41 (6.3 %)	164	150	81
2 NN, F	456	30 (6.6 %)	376	336	197
3 NN, F	501	19 (3.8 %)	209	186	122
4 NN, F	519	14 (2.7 %)	200	190	105
5 N, M	601	95 (15.8 %)	282	258	142
6 N, F	469	14 (3.0 %)	324	300	141
7 NN, F	770	157 (20.4 %)	219	208	87
8 NN, M	442	4 (0.9 %)	282	246	168
9 N, F	760	231 (30.4 %)	247	226	156
10 N, F	686	272 (39.7 %)	288	241	241
11 N, F	712	193 (27.1 %)	224	200	133
12 N, F	572	74 (12.9 %)	280	272	140
13 N, M	326	11 (3.4 %)	508	438	295
14 NN, F	619	51 (8.2 %)	265	248	112
15 NN, M	510	52 (10.2 %)	328	298	150
16 N, F	525	30 (5.7 %)	308	300	122
17 N, F	471	40 (8.5 %)	343	302	214
18 N, M	614	138 (22.5 %)	247	220	156
19 NN, F	670	110 (16.4 %)	234	218	111
20 NN, F	659	112 (17.0 %)	276	256	120

Table 1

Statistics for fixations inside the picture. NN refers to non-novice, N to novice, F to female and M to male.

SUPPLEMENT B TO "WHAT WE LOOK AT IN PAINTINGS: A COMPARISON BETWEEN EXPERIENCED AND INEXPERIENCED ART VIEWERS"

By Anna-Kaisa Ylitalo*, Aila Särkkä† and Peter Guttorp‡

University of Jyvaskyla, Chalmers University of Technology and University of Gothenburg†, University of Washington and Norwegian Computing Center* ‡

Statistics for saccades inside the picture.

Subject	mean saccade duration (ms)	median saccade duration (ms)	sd saccade duration (ms)	mean saccade length (px)	median saccade length (px)	sd saccade length (px)
1 NN	136	118	86	173	159	119
2 NN	55	44	50	143	128	78
3 NN	158	110	133	157	135	103
4 NN	141	102	131	137	133	76
5 N	74	54	71	141	123	97
6 N	74	50	64	190	176	117
7 NN	76	48	64	126	110	88
8 NN	119	102	73	173	142	114
9 N	83	56	92	125	108	104
10 N	84	47	107	122	98	109
11 N	121	50	162	114	88	90
12 N	90	46	108	181	167	107
13 N	80	72	45	208	199	100
14 NN	57	52	34	132	108	94
15 NN	69	44	88	169	156	102
16 N	55	44	47	138	117	87
17 N	82	46	90	113	92	87
18 N	122	62	126	180	157	130
19 NN	96	54	77	86	64	73
20 NN	61	46	52	191	179	99

Table 1

Statistics for saccades inside the picture. NN refers to non-novice, N to novice.

SUPPLEMENT C TO "WHAT WE LOOK AT IN PAINTINGS: A COMPARISON BETWEEN EXPERIENCED AND INEXPERIENCED ART VIEWERS"

By Anna-Kaisa Ylitalo*, Aila Särkkä† and Peter Guttorp‡

*University of Jyvaskyla**, *Chalmers University of Technology and University of Gothenburg†*, *University of Washington and Norwegian Computing Center* ‡

The most visited areas during the six 30 second intervals for novices and non-novices.

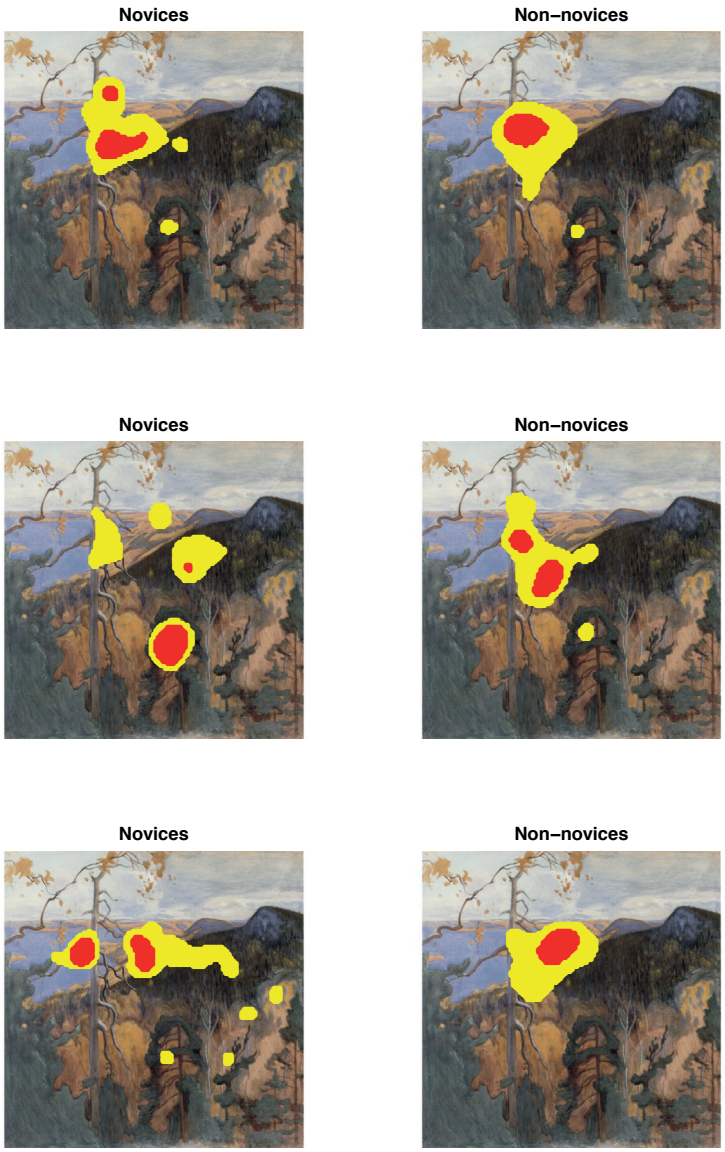

FIG 1. *Top 5 % (yellow) and top 1 % (red) intensities in 30 second intervals (0-30s), (30-60s), (60-90s), novices vs. non-novices.*

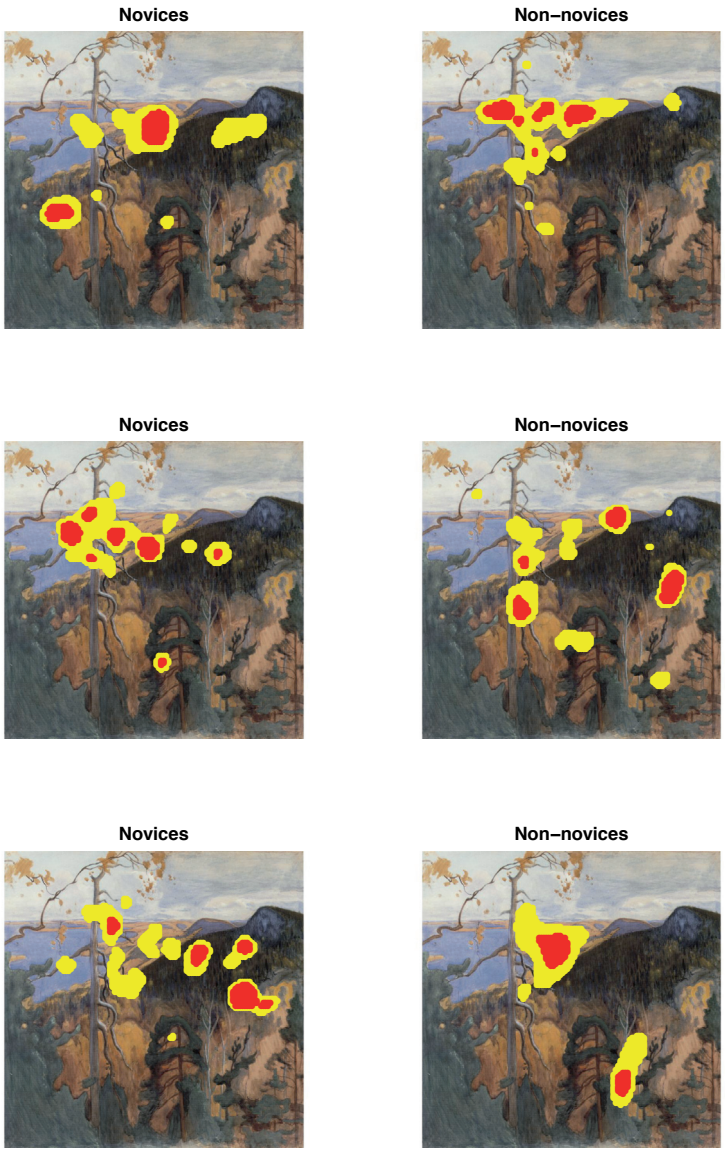

FIG 2. *Top 5 % (yellow) and top 1 % (red) intensities in 30 second intervals (90-120s), (120-150s), (150-180s), novices vs. non-novices.*

SUPPLEMENT D TO "WHAT WE LOOK AT IN PAINTINGS: A COMPARISON BETWEEN EXPERIENCED AND INEXPERIENCED ART VIEWERS"

By Anna-Kaisa Ylitalo*, Aila Särkkä† and Peter Guttorp‡

*University of Jyvaskyla**, *Chalmers University of Technology and University of Gothenburg*†, *University of Washington and Norwegian Computing Center* ‡

Results for the intensity surface and fixation duration distribution comparisons between novices and non-novices for all six paintings used in the experiment.

Painting	T_0	p-value
Järnefelt	59604	0.108
Monet	479771	0.590
Suomi	124453	0.768
Kandinsky	91837	0.078
Poussin	950777	0.683
Tammi	64840	0.675

Table 1

Results for the intensity surface comparisons.

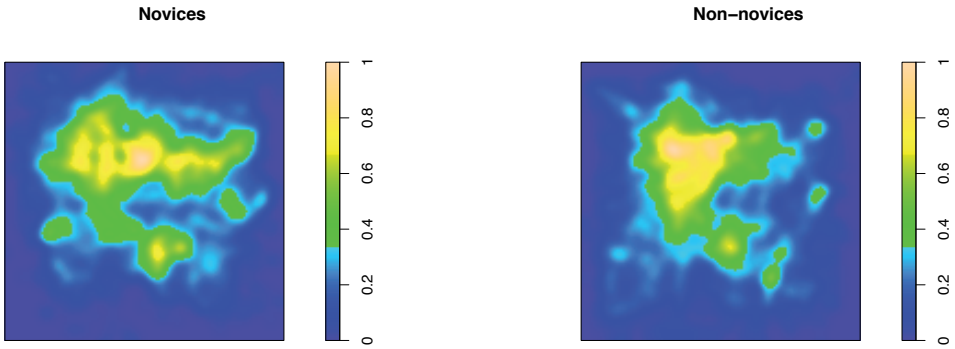

Fig 1. *The estimated intensity surfaces for the Koli-landscape (Järnefelt), bw = 17.*

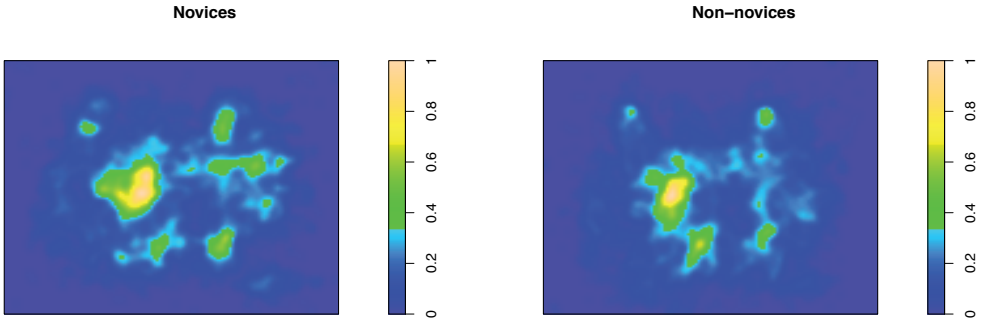

FIG 2. *The estimated intensity surfaces for the Terrasse à Sainte-Adresse (Monet), bw = 13.*

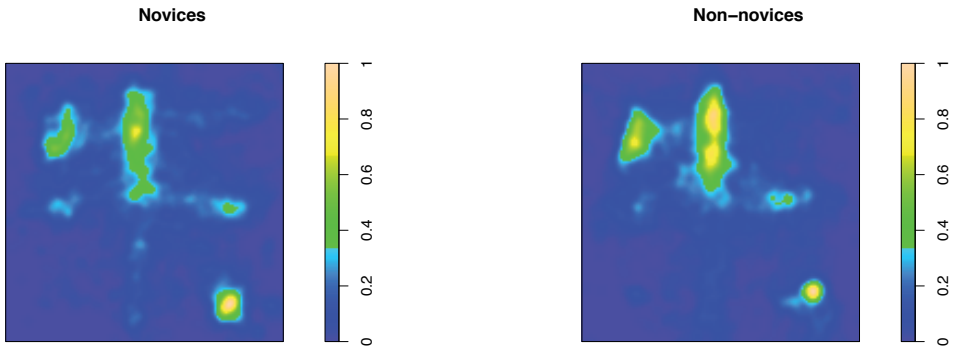

FIG 3. *The estimated intensity surfaces for La croisèe des destins (Suomi), bw = 12.*

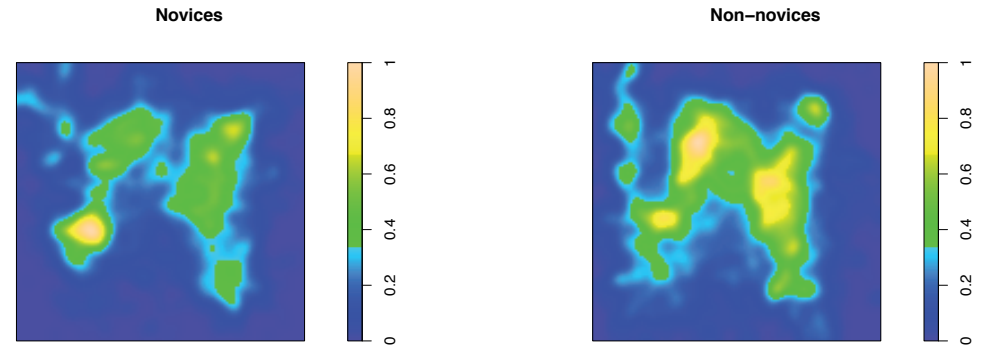

FIG 4. *The estimated intensity surfaces for the Picture with a Black Arch (Kandinsky), bw = 18.*

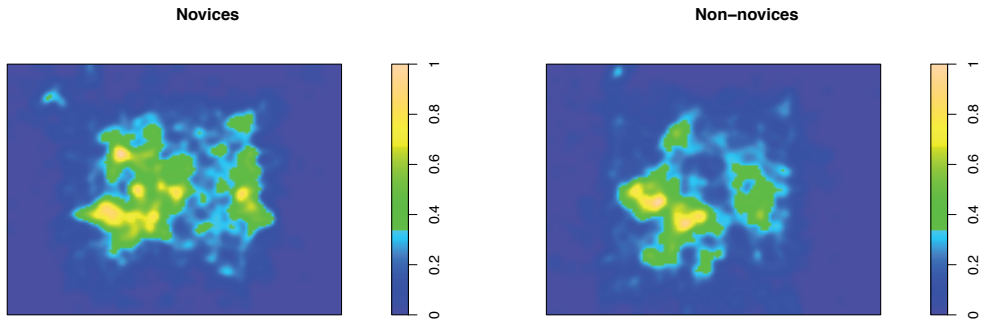

FIG 5. *The estimated intensity surfaces for the La Lamentation sur le Christ (Poussin), bw = 13.*

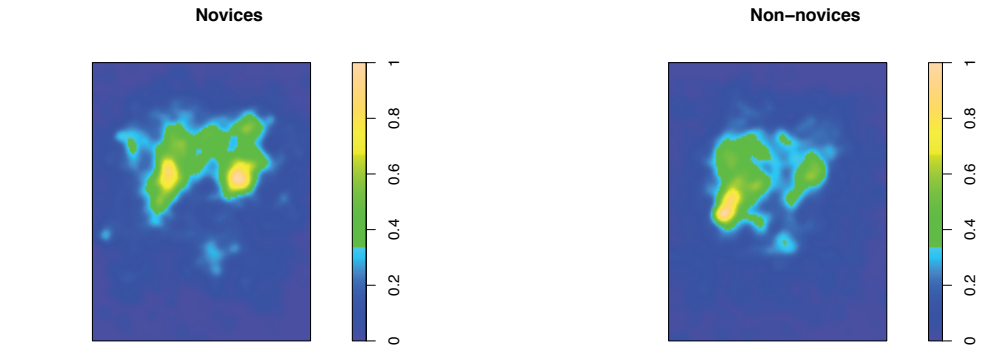

FIG 6. *The estimated intensity surfaces for the Poem Forces to Kneel Down (Tammi), bw = 13.*

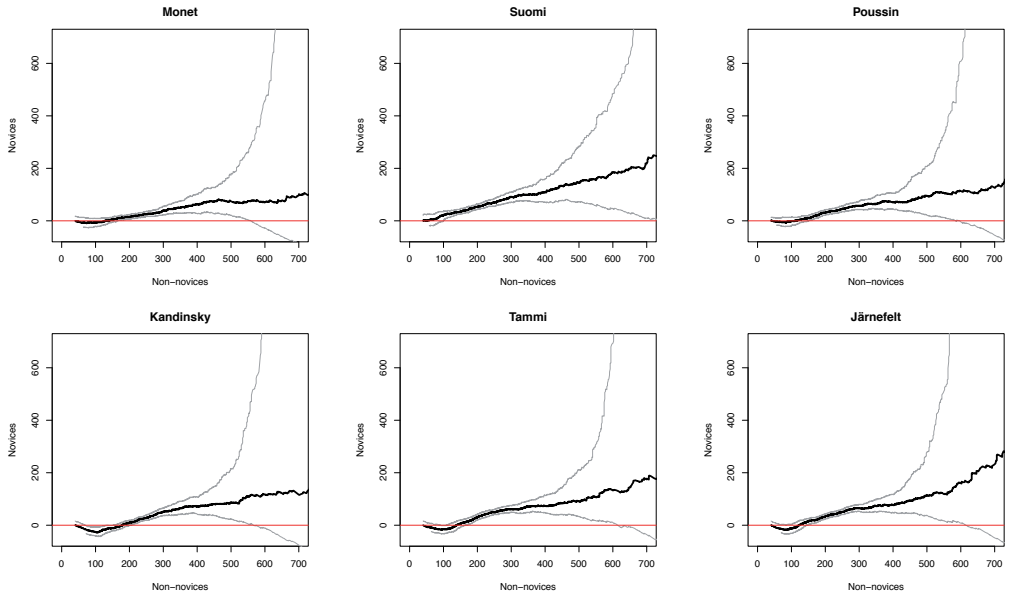

FIG 7. *Shift plots of fixation duration distributions for non-novices vs. novices and for each painting. The red line indicates equal distributions, and the grey lines form a simultaneous asymptotic 95% confidence band.*

SUPPLEMENT E TO "WHAT WE LOOK AT IN PAINTINGS: A COMPARISON BETWEEN EXPERIENCED AND INEXPERIENCED ART VIEWERS"

By Anna-Kaisa Ylitalo*, Aila Särkkä† and Peter Guttorp‡

*University of Jyvaskyla**, *Chalmers University of Technology and University of Gothenburg†*, *University of Washington and Norwegian Computing Center* ‡

Results for the fixation duration distribution comparisons within novice and non-novice groups for the three pairs of paintings: Järnefelt - Monet; Kandinsky - Suomi and Poussin - Tammi.

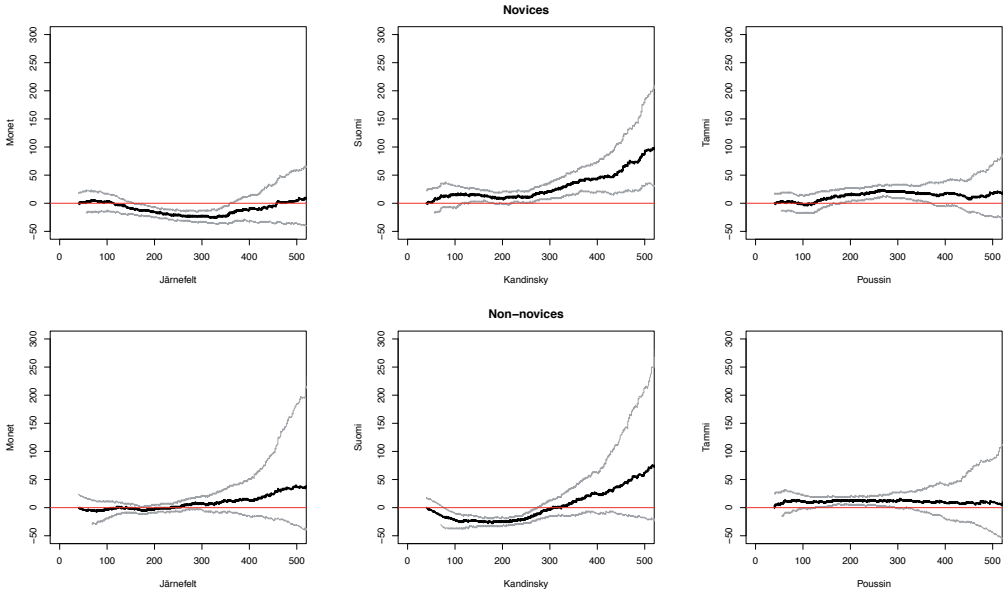

Fig 1. *Groupwise shift plots of the fixation duration distributions for the three pairs of paintings. The red line indicates equal distributions, and the grey lines form a simultaneous asymptotic 95% confidence band.*

III

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

Penttinen, A. and Ylitalo, A.-K.

Spatial Statistics, 17:1–21, 2016. doi:10.1016/j.spatsta.2016.03.005

Spatial Statistics 17 (2016) 1–21

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

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

Antti Penttinen [a], Anna-Kaisa Ylitalo [a,b,*]

[a] Department of Mathematics and Statistics, P.O. Box 35 (MaD), FI-40014 University of Jyvaskyla, Finland
[b] Department of Music, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland

ARTICLE INFO

Article history:
Received 30 October 2015
Accepted 28 March 2016
Available online 20 April 2016

Keywords:
Coverage
Heterogeneous media
Likelihood
Recurrence
Self-interacting random walk
Stochastic geometry

ABSTRACT

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.

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.

* Corresponding author at: Department of Music, P.O. Box 35, FI-40014 University of Jyvaskyla, Finland.
E-mail address: anna-kaisa.ylitalo@jyu.fi (A.-K. Ylitalo).

http://dx.doi.org/10.1016/j.spasta.2016.03.005
2211-6753/

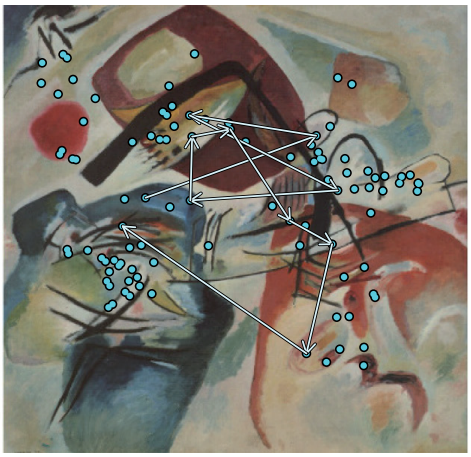

Fig. 1. First 100 fixation points of one subject on a painting called Picture with a Black Arch (or Black Bow) by Wassily Kandinsky (Kandinsky, 1912). The arrows show the movement of the gaze during the first three seconds.

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 Fig. 1.

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 et al. (2008) and Diggle (2013). Extensive software `spatstat` (Baddeley et al., 2015) has made efficient point pattern data analysis attractive. Point process statistics has been applied for eye movement data by Barthelmé et al. (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 et al. (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 et al. (2010), and an application to eye movement data by Ylitalo et al. (in press).

We consider eye movement data to be a realization of a sequential spatial point process which allows us to extend the approach by Barthelmé et al. (2013) for detecting new important dynamic structures of data. The advantage of this approach is that the likelihood is tractable and the simulation of realizations is straightforward. In addition, sequential point process modelling is a construction step for spatio-temporal point processes.

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

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

Our idea is that the history of the sequence changes the dynamics during the evolution of the eye movement sequence. We present two general principles for model construction, both of which are generalizations of the random walk in heterogeneous media. The first principle is the history-dependent thinning of transitions, which assigns smaller weights for suggested transitions being at odds with the chosen functional summary characteristic conditional on all previous fixations. This type of penalization is similar to the area-interaction process by Baddeley and van Lieshout (1995) in point process statistics. The second principle resembles the ARCH (autoregressive conditional heterogeneity) model, 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 $\overrightarrow{x}_n = (x_1, \ldots, x_n)$ is a sequence of time-ordered points in a bounded window $W \subset \mathbb{R}^2$. The corresponding unordered point set $\{x_1, \ldots, x_n\}$ is denoted by $\{\overrightarrow{x}_n\}$. 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 $\overrightarrow{x}_k = (x_1, \ldots, x_k)$, $k = 1, \ldots, (n-1)$, is denoted by $f_{k+1}(x|\overrightarrow{x}_k)$, a transition probability density for the transition $\overrightarrow{x}_k \to (\overrightarrow{x}_k, x)$, and the joint density of $\overrightarrow{x}_n$ is

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

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 $f_{k+1}(x|\overrightarrow{x}_k)$, $k = 1, \ldots, n-1$, 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

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

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

$$K(x_k, x) \geq 0 \quad \text{for all } x_k, x \in W, \quad \text{and}$$

$$\int_W K(x_k, u)\, \mathrm{d}u = 1 \quad \text{for all } x_k \in W,$$

$k = 1, \ldots, n-1$. 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 et al., 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

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

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 $\overrightarrow{x}_k \to (\overrightarrow{x}_k, x)$ is assumed to be

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

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

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 $\{\overrightarrow{x}_k\}$, denoted by $\mathrm{Conv}(\overrightarrow{x}_k)$, is the minimal convex subset of W which contains all the points of $\{\overrightarrow{x}_k\}$. 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 $S_C(\overrightarrow{x}_k)$ the area of $\mathrm{Conv}(\overrightarrow{x}_k)$ and call it convex hull coverage.

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

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

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 $S_B(\overrightarrow{x}_k)$ 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 $\mathrm{Conv}(\overrightarrow{x}_k)$, the odds ratio of acceptance w.r.t. a proposal $y \in \mathrm{Conv}(\overrightarrow{x}_k)$ with $\|x - x_k\| = \|y - x_k\|$ is $S_C(\overrightarrow{x}_k, x)/S_C(\overrightarrow{x}_k)$.

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

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

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 $\overrightarrow{x}_k \to (\overrightarrow{x}_k, x)$ with the truncated Gaussian kernel is

$$f_{k+1}(x|\overrightarrow{x}_k) \propto \alpha(x)\, e^{-\frac{1}{2\sigma^2}\|x_k - x\|^2} \left(\mathbf{1}_{W \setminus \mathrm{Conv}(\overrightarrow{x}_k)}(x) + \rho\, \mathbf{1}_{\mathrm{Conv}(\overrightarrow{x}_k)}(x)\right), \tag{6}$$

where $k = 1, 2, \ldots, n-1$, 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 $\mathrm{Conv}(\overrightarrow{x}_k)$ is replaced by $\mathrm{Bcov}(\overrightarrow{x}_k)$ 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 $\tilde{S}_R(\overrightarrow{x}_k, x) = \sum_{i=1}^{k} \mathbf{1}_{b(x_i,r)}(x)$. However, instead of using all the previous points, at step k the number of earlier visits is calculated from the point set $\{\overrightarrow{x}_{k-1}\}$ omitting the most recent point x_k. This delayed recurrence measure

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

is less confounded with the Markovian kernel $K(x_k, x)$ than $\tilde{S}_R(\overrightarrow{x}_k, x)$ 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

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

where $\theta \in [0, 1]$ and $k \geq 2$. The odds ratio for accepting a location close to the points of $\{\overrightarrow{x}_{k-1}\}$ 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 $\{\overrightarrow{x}_{k-1}\}$.

If $\theta = 0.5$, the process is a random walk. The density for the transition $\overrightarrow{x}_k \to (\overrightarrow{x}_k, x)$ with the truncated Gaussian kernel is

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

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

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 $\{\overrightarrow{x}_k\}$ 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

$$f_{k+1}(x|\overrightarrow{x}_k) = \frac{\alpha(x)\, K_{\phi_k}(x_k, x)}{\int_W \alpha(u)\, K_{\phi_k}(x_k, u)\mathrm{d}u} \tag{10}$$

$$\phi_k = \phi_k(\overrightarrow{x}_k) \propto H(S(\overrightarrow{x}_k)) \tag{11}$$

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

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

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

$\tau, \kappa \geq 0$ and $x_k, x \in W$, where $S(\overrightarrow{x}_k)$ is the coverage of $\{\overrightarrow{x}_k\}$. 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 $\overrightarrow{x}_n = (x_1, \ldots, x_n)$ 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

$$l(\sigma^2, \rho) = \sum_{k=1}^{n-1} \log(\alpha(x_{k+1})) - \frac{1}{2\sigma^2} \sum_{k=1}^{n-1} \|x_k - x_{k+1}\|^2 + \log(\rho) \sum_{k=1}^{n-1} 1_{\mathrm{Conv}(\overrightarrow{x}_k)}(x_{k+1}) - \sum_{k=1}^{n-1} \log \int_W \alpha(u)\, e^{-\frac{1}{2\sigma^2}\|x_k-u\|^2} \left(1_{W\setminus \mathrm{Conv}(\overrightarrow{x}_k)}(u) + \rho\, 1_{\mathrm{Conv}(\overrightarrow{x}_k)}(u)\right) \mathrm{d}u \tag{14}$$

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

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

where the expectation is over the conditional distribution of a new random point U from the distribution $f_{k+1}(x|\overrightarrow{x}_k)$. 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 a simple choice also for more complex model.

History-adapted model

The kernel width ϕ_k of the kth 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 $S(\overrightarrow{x}_k)$. The log-likelihood for the general model given by (10) and (11) is

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

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

$$l(\tau, \kappa) = \sum_{k=2}^{n-1} \log(\alpha(x_{k+1})) - \sum_{k=2}^{n-1} \frac{1}{2\phi(\tau,\kappa)} \|x_k - x_{k+1}\|^2 - \sum_{k=2}^{n-1} \log \int_W \alpha(u) e^{-\frac{1}{2\phi(\tau,\kappa)}\|x_k-u\|^2} \mathrm{d}u \tag{15}$$

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

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

$$\sum_{k=2}^{n-1} \mathbf{E}_{\tau,\kappa} \left(\|x_k - U\|^2 \mid \overrightarrow{x}_k \right) / \phi_k = \sum_{k=2}^{n-1} \|x_k - x_{k+1}\|^2 / \phi_k$$

$$\sum_{k=2}^{n-1} \mathbf{E}_{\tau,\kappa} \left(S(\overrightarrow{x}_k)\, \|x_k - U\|^2 \mid \overrightarrow{x}_k \right) / \phi_k = \sum_{k=2}^{n-1} S(\overrightarrow{x}_k)\, \|x_k - x_{k+1}\|^2 / \phi_k,$$

where the expectations are over the conditional distribution $f_{k+1}(x|\overrightarrow{x}_k)$ 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 et al. (2011) and in Baddeley et al. (2014).

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

3. Simulation experiments

Realizations 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 $(x_1,\ldots,x_k)$ are simulated. A simple algorithm for adding a point x to $\overrightarrow{x}_k = (x_1,\ldots,x_k)$, or equivalently, simulating from the distribution having density $f_{k+1}(x|\overrightarrow{x}_k)$, 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 $\pi(x, S(\overrightarrow{x}_k, x))$.
- History-adapted model: The kernel width $\phi_k = H(S(\overrightarrow{x}_k))$ 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 realizations 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 realization 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 realizations. 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.

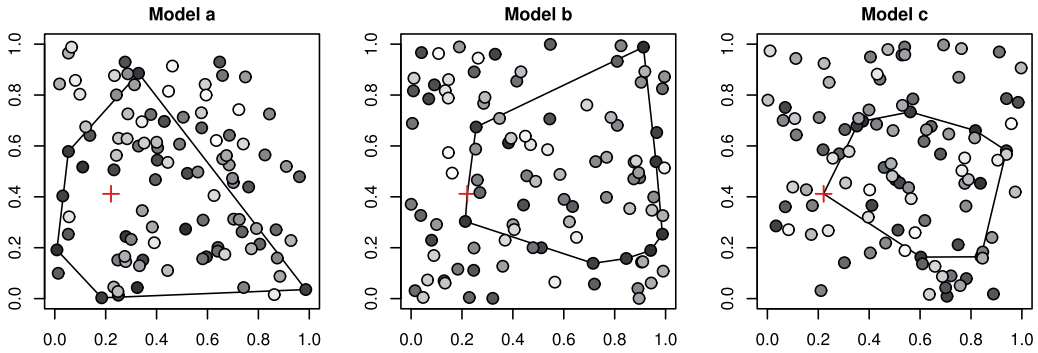

Fig. 2. Simulated patterns of the three history-dependent models with coverage self-interaction. The colour of the points denotes their order (from dark to light) and the first fixation is marked with a cross. The polygon indicates the convex hull of the first 10 points.

We simulate 19 realizations of the random walk model *a*, since it here represents a reference model, and five realizations of Model *b* and Model *c*. One of the simulated realizations of each model can be seen in Fig. 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.

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 Fig. 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.

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 realizations of the random walk model *d* as well as five realizations of Model *e* and Model *f*. In Fig. 4 we can see that, compared with the random walk model *d*, the realization of the low recurrence model *e* indicates a tendency towards higher regularity, and the realization 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.

The results of the functional summary statistics are depicted in Fig. 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

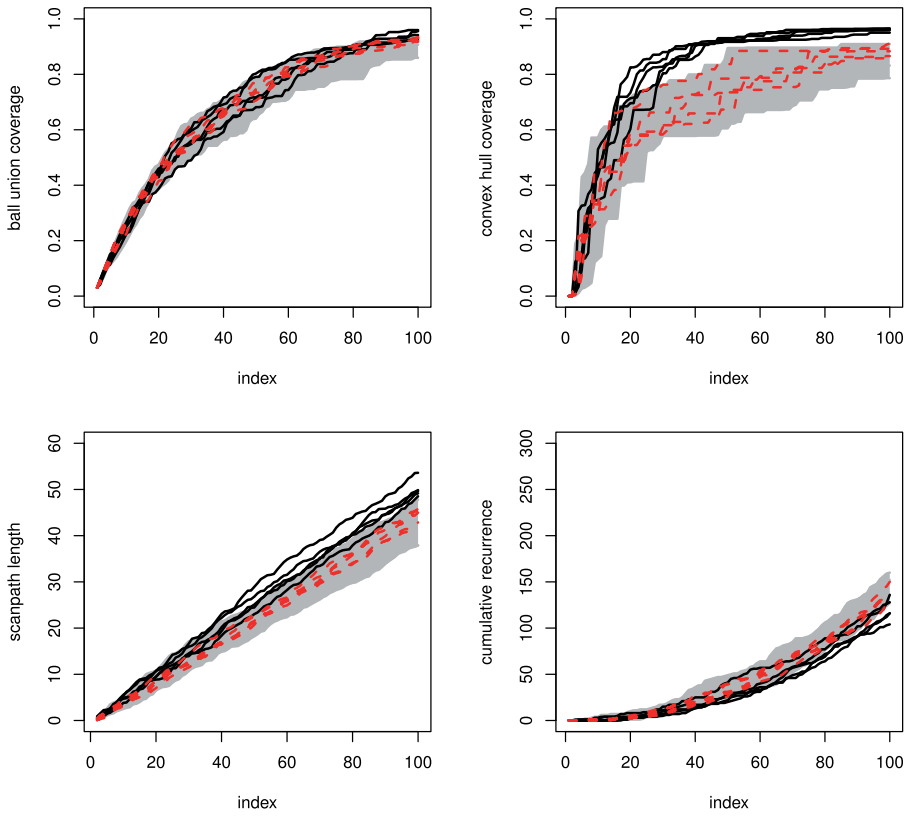

Fig. 3. Ball union coverage with radius 0.1 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 0.1 (bottom right) for the two models (five realizations of each): The fast coverage model *b* is marked with black solid lines and the mild coverage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

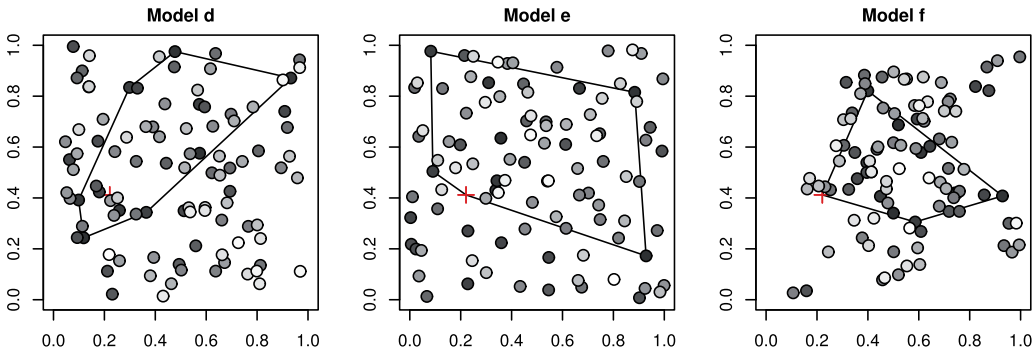

Fig. 4. Simulated patterns of the three history-dependent models with recurrence self-interaction. The colour of the points denotes their order (from dark to light) and the first fixation is marked with a cross. The polygon indicates the convex hull of the first 10 points.

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 scanpath 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.

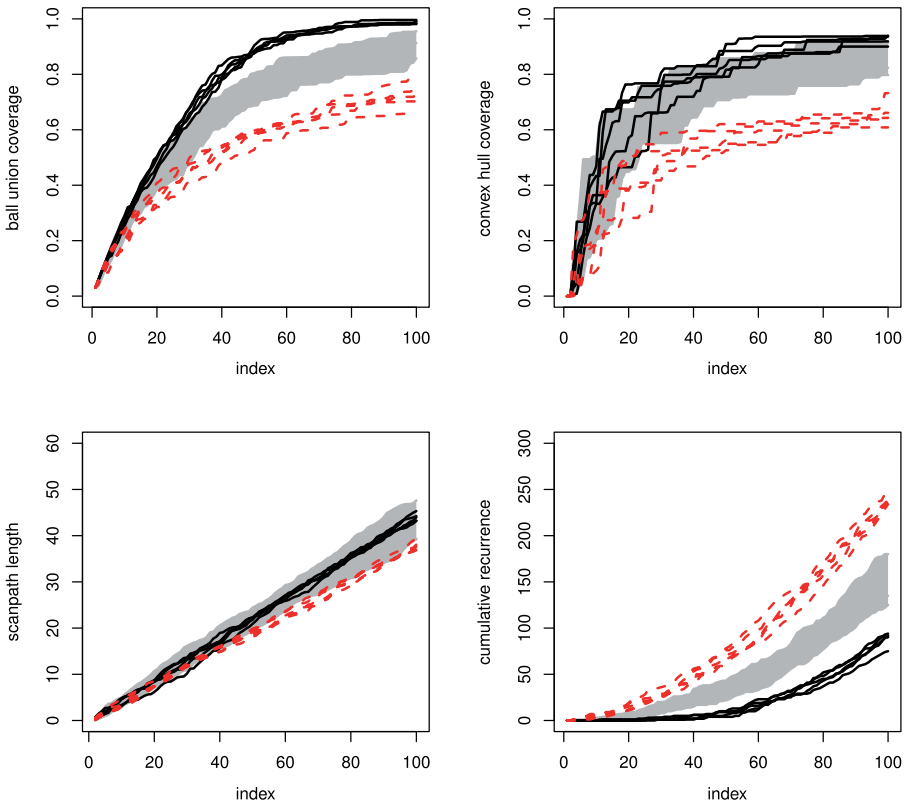

Fig. 5. Ball union coverage with radius 0.1 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 0.1 (bottom right) for the two models (five realizations of each): The low recurrence model e is marked with black solid line and the high recurrence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

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 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) $S(\overrightarrow{x}_k)$ is the area of the convex hull coverage generated by the points $(x_1, \ldots, x_k)$. 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 Figs. 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 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.

The functional summary statistics are plotted in Fig. 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

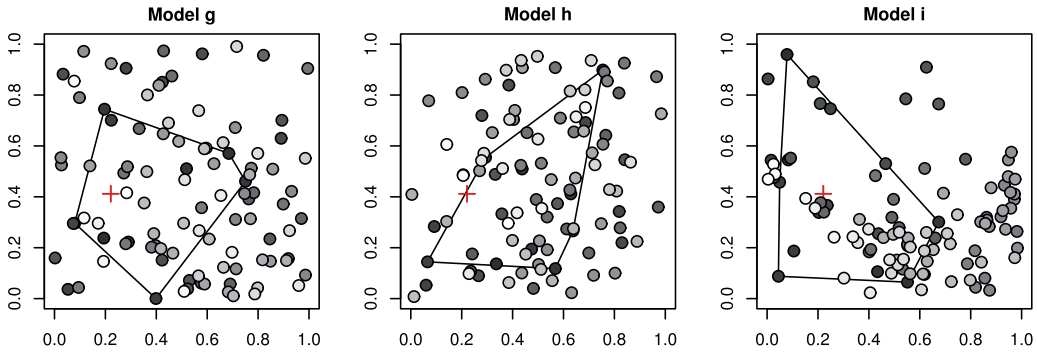

Fig. 6. Simulated patterns of the three history-adapted models. The colour of the points denotes their order (from dark to light) and the first fixation is marked with a cross. The polygon indicates the convex hull of the first 10 points.

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: Picture with a Black Arch 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 X^{TM}Hi-Speed eye tracker with temporal resolution of 500 Hz. The distance between a participant's head and the screen was about 85 cm, and a forehead rest was used in order to prevent 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 A Picture with a Black Arch (1912) by Wassily Kandinsky shown in Fig. 1. 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 s time-interval, as shown in Fig. 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 Kandinsky'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

Table 1
Components of the four rejection models.

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

* The value of σ^2 should be chosen to be large enough such that the kernel is flat in the specified window.

presented in Fig. 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 Fig. 8. This particular subject paid most attention to the areas with high intensity, but the gaze stayed in some low intensity areas also.

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 $S_R(\overrightarrow{x}_k, x)$, $k = 2, \ldots, n-1$ 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 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

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

The likelihood for Model 1 is just the first term on the right hand side of Eq. (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 $(60, 80, \ldots, 400)$ and $(0.05, 0.10, \ldots, 0.95)$ 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.

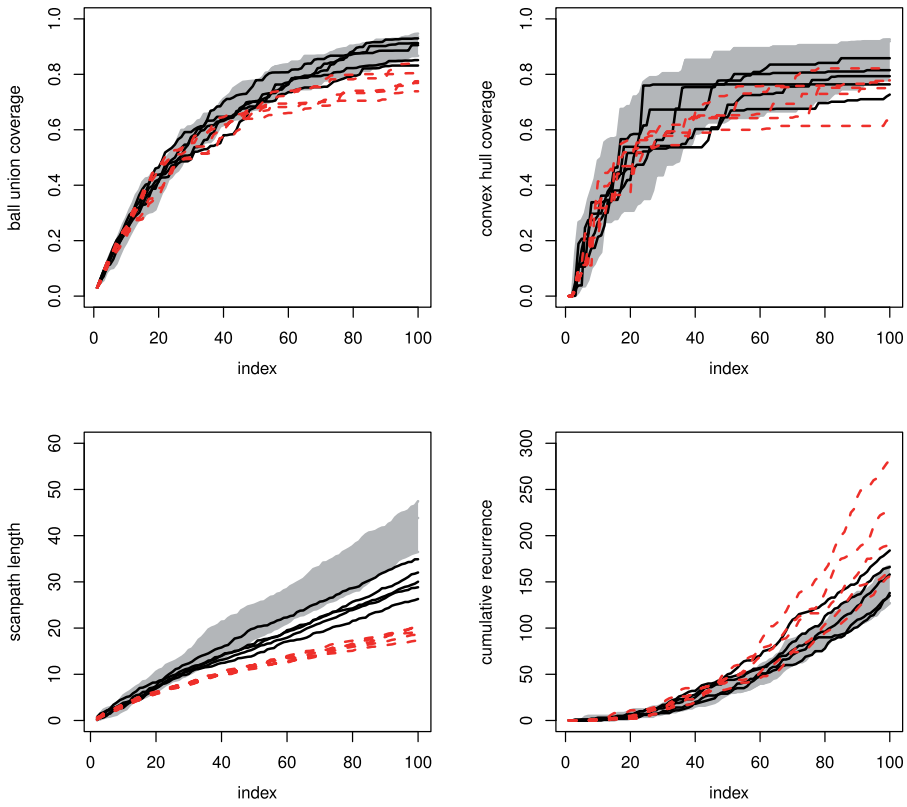

Fig. 7. Ball union coverage with radius 0.1 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 0.1 (bottom right) for the two models (five realizations of each): The mild clustering model h is marked with black solid lines and the fast clustering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

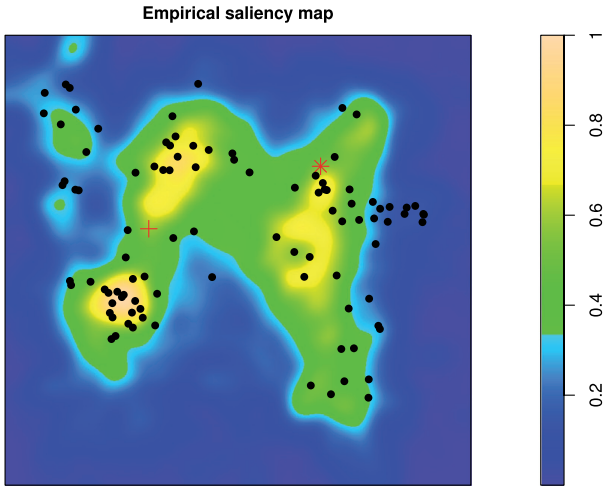

Fig. 8. Empirical saliency map estimated from the fixations of all subjects excluding the one under study. The points indicate the first 100 fixation points of the particular subject under study. The first fixation is marked with a cross and the second with a star.

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 realizations 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 realization of each model can be seen in

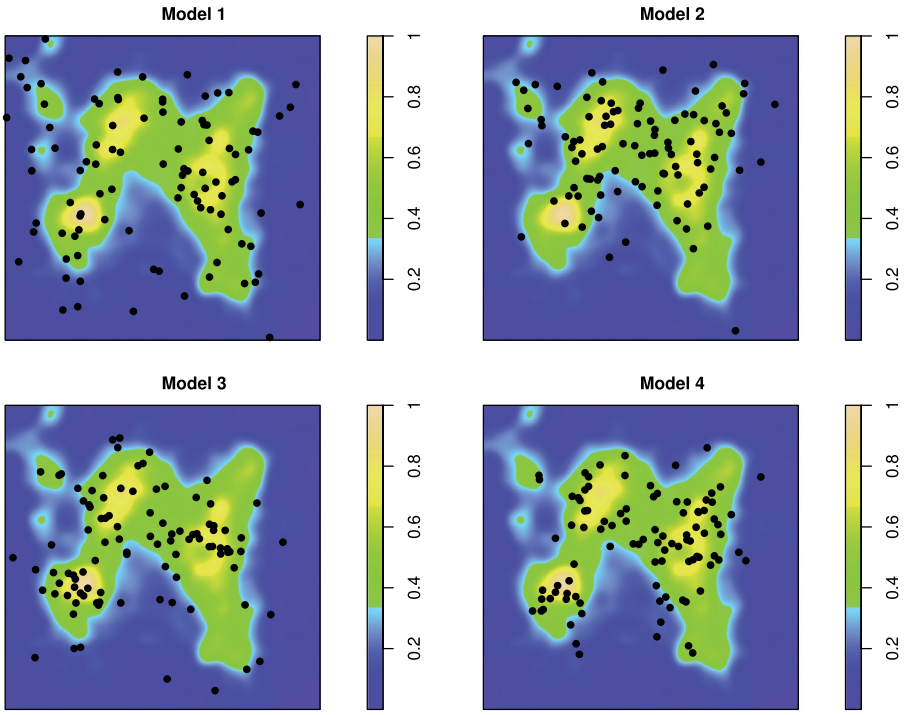

Fig. 9. Simulated realizations of the four models fitted to the eye movement data of the particular subject under study, overlaid with the empirical saliency map based on data from the other subjects. The first two points are fixed and marked with a cross and a star, respectively.

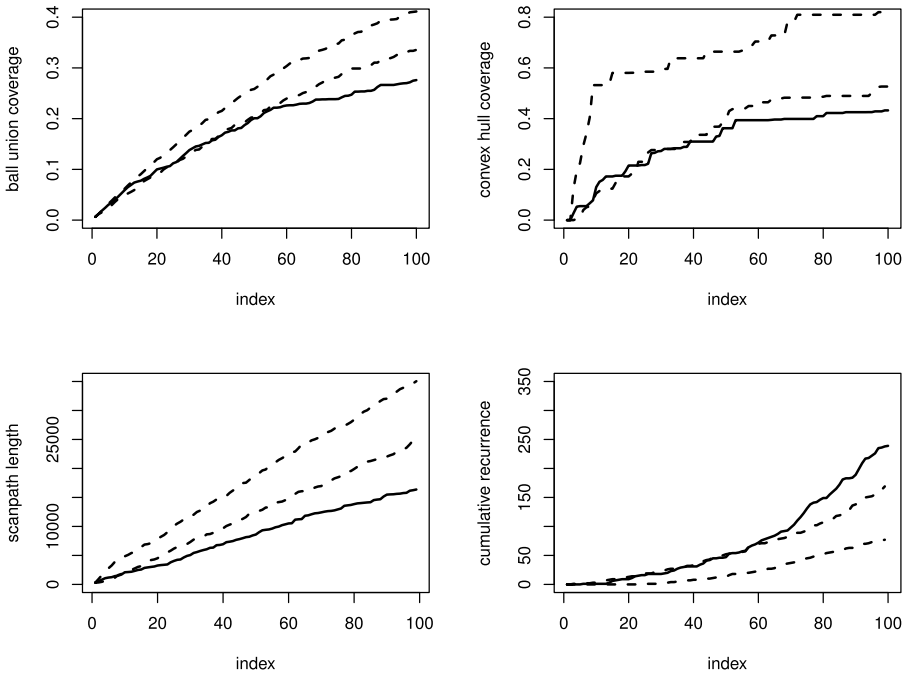

Fig. 10. Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (solid line). Dashed lines represent pointwise envelopes estimated from 99 simulations of Model 1.

Fig. 9, and the summary statistics estimated from the data with pointwise envelopes estimated from the simulations in Figs. 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].

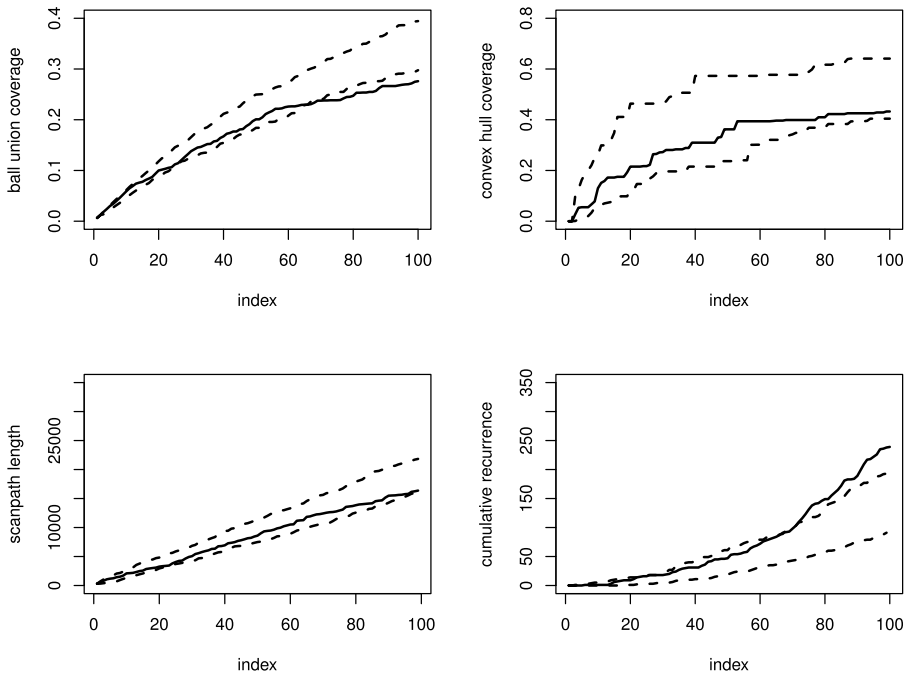

Fig. 11. Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (solid line). Dashed lines represent pointwise envelopes estimated from 99 simulations of Model 2.

For Model 1 all summary statistics show poor fit (Fig. 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 Fig. 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 (Fig. 11). These findings indicate that data begin to cluster at the end of the inspection period, but the model does not carry that effect.

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 (Fig. 12). The marginal spatial structure looks slightly more clustered than Model 1 and Model 2 predict (Fig. 9).

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 Fig. 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 Fig. 13, the envelopes of Model 4 seem to cover the convex hull and ball union coverage

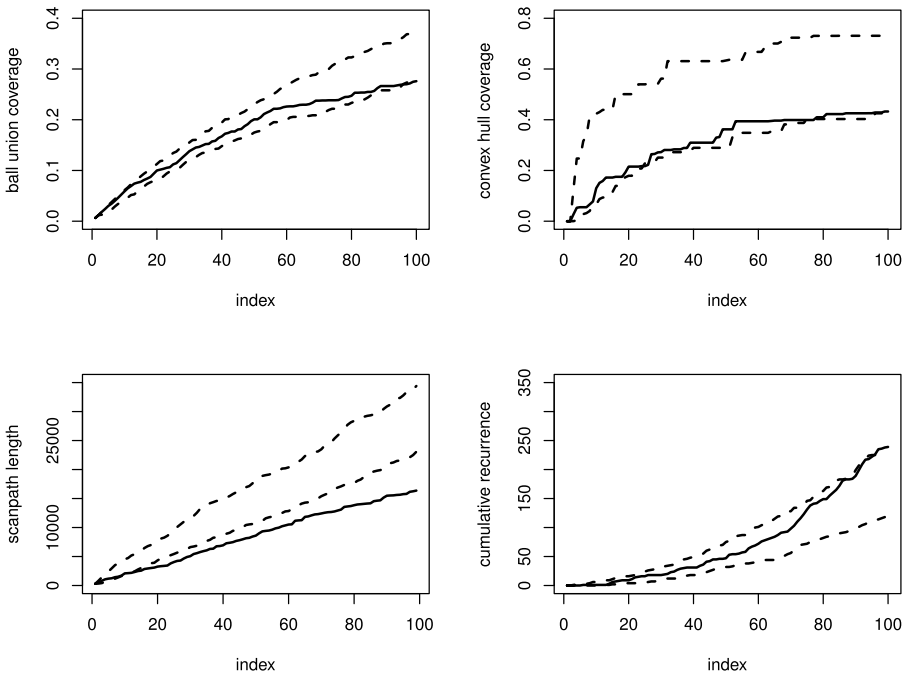

Fig. 12. Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (solid line). Dashed lines represent pointwise envelopes estimated from 99 simulations of Model 3.

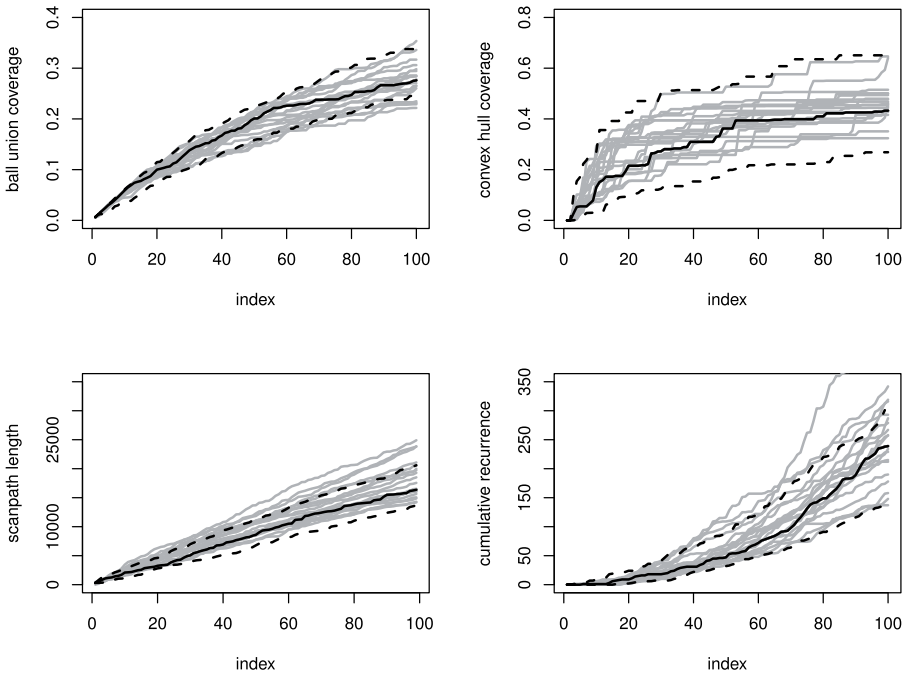

Fig. 13. Ball union coverage with radius 35 (top left), convex hull coverage (top right), scanpath length (bottom left) and cumulative recurrence with radius 50 (bottom right) for the subject under study (black solid line) and for all other subjects (grey solid lines). Dashed lines represent pointwise 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

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).

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

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 realizations 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 Fig. 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.

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.

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 $K(\overrightarrow{x}_k, x)$ is replaced by

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

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 $p(\overrightarrow{x}_k)$ 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 originates mainly from the target common to all participants and to the case under study and measures the wanted heterogeneity.

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) and 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 experiment 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, the authors 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).

References

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. Ecol. Monograph 84 (3), 477–489.

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

Baddeley, A.J., van Lieshout, M.N.M., 1995. Area-interaction point processes. Ann. Inst. Statist. Math. 47 (4), 601–619.

Barthelmé, S., Trukenbrod, H., Engbert, R., Wichmann, F., 2013. Modelling fixation locations using spatial point processes. J. Vision 13 (12), 1–34.

Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. CRC Press, Boca Raton.

Diggle, P.J., Gómez-Rubio, V., Brown, P.E., Chetwynd, A.G., Gooding, S., 2007. Second-order analysis of inhomogeneous spatial point processes using case–control data. Biometrics 63 (2), 550–557.

Diggle, P.J., Kaimi, I., Abellana, R., 2010. Partial-likelihood analysis of spatio-temporal point-process data. Biometrics 66 (2), 347–354.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. Chapman and Hall/CRC Press.

Engbert, R., Trukenbrod, H.A., Barthelmé, S., Wichmann, F.A., 2015. Spatial statistics and attentional dynamics in scene viewing. J. Vision 15 (1), 1–17.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50 (4), 987–1007.

Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood. Interface Foundation of North America. Retrieved from the University of Minnesota Digital Conservancy, http://purl.umn.edu/58440.

Grabarnik, P., Myllymäki, M., Stoyan, D., 2011. Correct testing of mark independence for marked point patterns. Ecol. Model. 222 (23), 3888–3894.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. Wiley, Chichester.
Itti, L., Koch, C., 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40, 1489–1506.
Kandinsky, W., 1912. Picture with a Black Arch. Musée National d'Art Moderne, Centre Georges Pompidou, Paris, France., Wikiart. Web. 6 Apr. 2016. http://www.wikiart.org/.
Kümmerer, M., Wallis, T., Bethge, M., 2014. How close are we to understanding image-based saliency? arXiv preprint arXiv:1409.7686.
Locher, P.J., 2006. The usefulness of eye movement recordings to subject an aesthetic episode with visual art to empirical scrutiny. Psychol. Sci. 48, 106–114.
Locher, P.J., Gray, S., Nodine, C.F., 1996. The structural framework of pictorial balance. Perception 25 (12), 1419–1436.
Locher, P.J., Krupinski, E.A., Mello-Thoms, C., Nodine, C.F., 2007. Visual interest in pictorial art during an aesthetic experience. Spatial Vis. 21, 55–77.
Rayner, K., 2009. Eye movements and attention in reading, scene perception, and visual search. Q. J. Exp. Psychol. 62, 1457–1506.
Ripley, B.D., 1987. Stochastic Simulation. Wiley.
Tatler, B.W., Baddeley, R.J., Vincent, B.T., 2006. The long and the short of it: Spatial statistics at fixation vary with saccade amplitude and task. Vis. Res. 46 (12), 1857–1862.
van Lieshout, M.N.M., 2006a. Campbell and moment measures for finite sequential spatial processes. Probab. Netw. Algorithms 1, 1–10. Report.
van Lieshout, M.N.M., 2006b. Maximum likelihood estimation for random sequantial adsorption. Adv. Appl. Probab. 38 (SGSA), 889–898.
van Lieshout, M.N.M., 2009. Sequential spatial processes for image analysis. In: Stereology and Image Analysis. Ecs10. Proceedings of the 10th European Congress of ISS. Bologna, Italy.
Ylitalo, A.-K., Särkkä, A., Guttorp, P., What we look at in paintings: A comparison between experienced and inexperienced art viewers, Ann. Appl. Stat. (in press).