ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Applied Statistics
Series C

Appl. Statist. (2017)

Markov models for ocular fixation locations in the presence and absence of colour

Adam B. Kashlak, Eoin Devane, Helge Dietert and Henry Jackson

University of Cambridge, UK

[Received February 2016. Revised February 2017]

Summary. In response to the 2015 Royal Statistical Society's statistical analytics challenge, we propose to model the fixation locations of the human eye when observing a still image by a Markov point process in $\mathbb{R}^2$. Our approach is data driven using k-means clustering of the fixation locations to identify distinct salient regions of the image, which in turn correspond to the states of our Markov chain. Bayes factors are computed as the model selection criterion to determine the number of clusters. Furthermore, we demonstrate that the behaviour of the human eye differs from this model when colour information is removed from the given image.

Keywords: Bayesian model selection; Cluster analysis; Finite mixture model; Image saliency; Markov point process; Ocular fixation

1. Introduction

Ocular movement data have posed a particularly tough challenge to researchers and offer many potential insights into human visual behaviour as well as many practical applications. As the most detailed information about a visual scene confronting the eye can only be extracted through the relatively small fovea at the retina's centre, complex ocular movements have evolved to absorb quickly as much information as possible through the eye (Zuber, 1981; Hacisalihzade *et al.*, 1992). The contribution, if any, of colour information to vision through saliency models and fixation location prediction has been heavily investigated (Baddeley and Tatler, 2006; Frey *et al.*, 2008; Ho-Phuoc *et al.*, 2012; Hamel *et al.*, 2014; Amano and Foster, 2014). Much research has gone into understanding ocular movement from the rapidly jerking saccades to the relatively still fixations. A better understanding of such movements has a wide range of applications from diverse fields of research such as evolutionary biology (Dominy and Lucas, 2001; Sumner and Mollon, 2000), neuroscience (Koch and Ullman, 1987; Desimone and Duncan, 1995), image segmentation (Ko and Nam, 2006; Achanta *et al.*, 2008) and image compression and resizing (Chen *et al.*, 2003; Wang *et al.*, 2003; Avidan and Shamir, 2007).

In this paper, we shall specifically focus on the eye's fixations by modelling such a sequence of fixations as a point process in $\mathbb{R}^2$. The distribution of fixations over a given image is treated as a finite mixture model comprised of disjoint *salient* regions, which correspond to the interesting bits of the image; see McLachlan and Peel (2004). This set of salient regions is used as the state space of a Markov chain. Each fixation is then an observation from the mixture component corresponding to the current state of the Markov chain. Under this model, it is shown that the

Address for correspondence: Adam B. Kashlak, Cambridge Centre for Analysis, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WA, UK.
E-mail: ak852@cam.ac.uk

 0035–9254/17/67000
Published by John Wiley & Sons Ltd on behalf of the Royal Statistical Society.

presence or absence of colour information in the image drastically affects the behaviour of a given sequence of fixations.

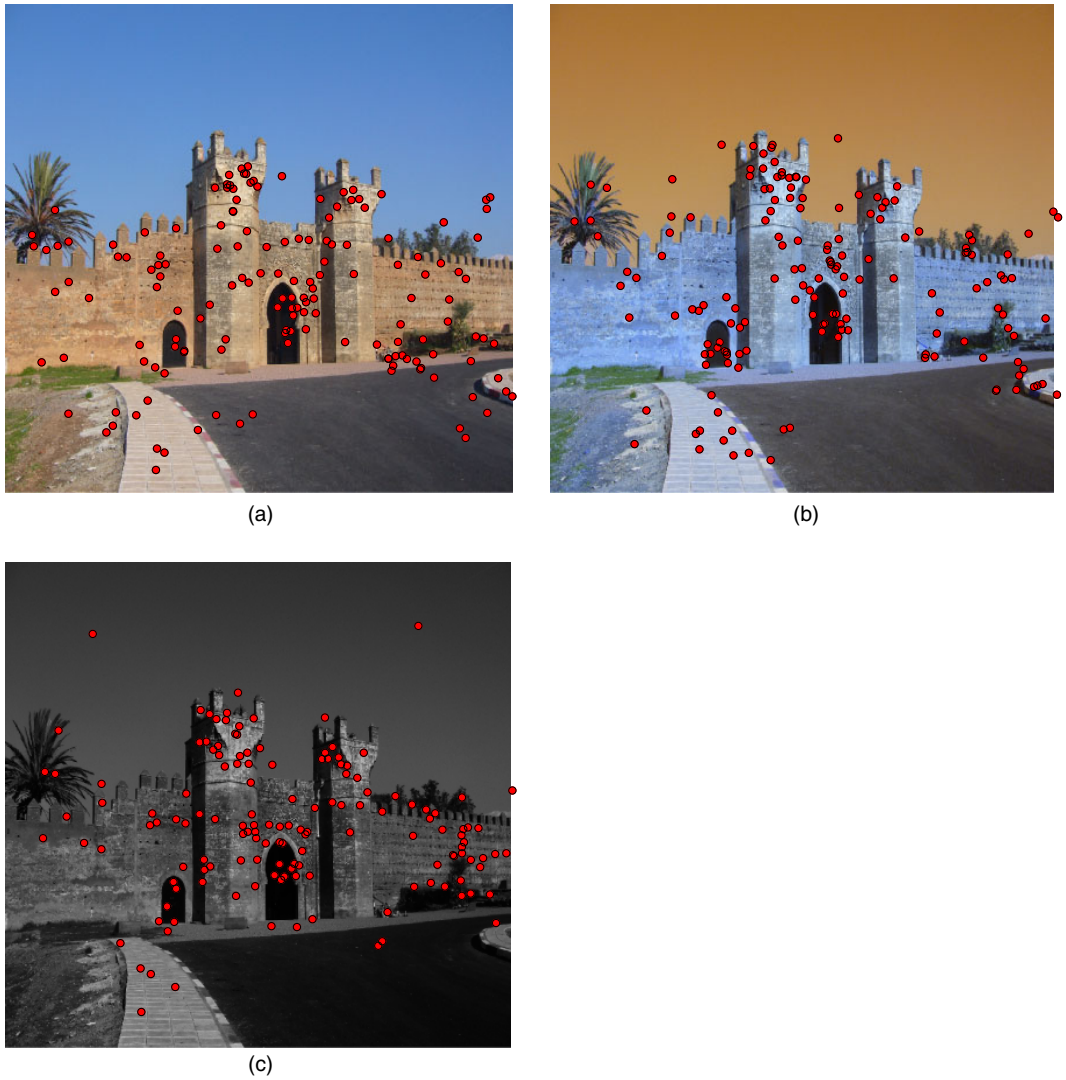
There has been much past research on the statistical analysis of static spatial point patterns and spatial point processes; for an overview, see Diggle (2003) and Illian *et al.* (2008). The inhomogeneous Poisson process and its generalizations, such as the Cox process (Cox and Isham, 1980; Møller *et al.*, 1998; Brix and Diggle, 2001; Diggle *et al.*, 2013; Taylor *et al.*, 2013) in which the intensity of the process is itself a stochastic process, have demonstrated tremendous value to inference in the realm of spatial statistics. However, these approaches treat the pairs of points as independent conditional on such an underlying stochastic process. Alternatively, the incorporation and estimation of an interaction term for such Markov point processes has been also investigated (Baddeley *et al.*, 2000; Jensen and Nielsen, 2000; Berthelsen and Møller, 2008). These interactions generally take on the form of preferences for attraction or repulsion between points. In contrast, our approach is a little more heavy handed by cutting the plane into a finite set of states for a Markov chain. The argument for taking this approach arises from the intuition that most still images are comprised of a finite number of interesting objects to examine, which will be our Markov states. This idea is similar to the earlier work of Stark and Ellis (1981) and Hacisalihzade *et al.* (1992). However, whereas Hacisalihzade *et al.* (1992) manually partitioned an image, we shall take a data-driven approach to segment an image into a set of finite disjoint pieces of interest.

It is worth emphasizing that these 'finite disjoint pieces of interest' are constructed solely from the data. They may correspond to regions of high saliency, which are those regions of sharp local contrast (Koch and Ullman, 1987; Itti and Koch, 2000). These are the regions whose 'features differ from the surrounding features' (Ho-Phuoc *et al.*, 2012) or are surprising in an information theoretic sense (Bruce and Tsotsos, 2009). However, saliency lacks a concrete definition and is often thought of as an amalgamation of low level image features such as sharp contrasts in luminosity and colour channels. Some researchers (Tatler *et al.*, 2011; Schütz *et al.*, 2011) have taken a more critical stance towards saliency stating that it ignores such high level image features as faces and text (Cerf *et al.*, 2009), which particularly attract the eye's gaze. As well, such low level features do not take into account the learned behavioural aspects of the eye's movements, systematic tendencies between successive fixations and viewing bias (Tatler and Vincent, 2008; Amano and Foster, 2014; Le Meur and Coutrot, 2016). Our analysis allows us to remain agnostic towards such debates. The data define the interesting regions, and it is left to vision researchers to decide, for example, whether a given region is interesting because it contains high level semantic information or low level luminosity contrast or some other feature which draws the gaze of the eye.

Modelling eye fixations as a spatial point pattern was previously discussed in Barthelmé *et al.* (2013) where an inhomogeneous Poisson process was utilized. The location-dependent rate parameter of the inhomogeneous Poisson process was determined by a measure of the saliency of each region of a given photograph. Alternatively, Kümmerer *et al.* (2014) applied deep neural networks to identify salient regions and ultimately to predict fixations. But, as is mentioned in Ho-Phuoc *et al.* (2012),

> 'there is no computational saliency model that can predict an observer's fixation location better than the model using fixations from other subjects'.

In light of that, we take a data-driven approach to modelling sequential fixation locations by using nine of the 10 subjects to train our model and the 10th for validation. Bayes factors are used as a model selection criterion; see Good (1967) for the use of Bayes factors in the multinomial hypothesis setting, and Kass and Raftery (1995) for a general overview of Bayes factors and



(a) (b)

(c)

Fig. 1. Example of the three colour schemes, (a) normal, (b) abnormal and (c) greyscale, under analysis with plotted fixations of all 10 subjects on each

model selection. Our modelling approach does not incorporate any direct information about the images themselves. With additional thought, our Markov states could ultimately be constructed from a saliency map and the high level information in the images rather than from the data.

The data under scrutiny come from the study of Ho-Phuoc *et al.* (2012) who were interested in whether the presence, absence or modification of the colour in a given image affects how the eye moves when looking at this image. Their study consisted of three colour schemes: normal colours, abnormal colours and greyscale. Normal refers to the unmodified image. Abnormal corresponds to swapping the red–green and blue–yellow chrominance channels. Greyscale corresponds to the complete removal of all colour information. An example of the three colour schemes with plotted fixations for all 10 subjects is displayed in Fig. 1.

The data were collected as follows. 10 observers were selected for each of the three colour schemes totalling 30 subjects in all. Each subject was presented with 60 photographs under a fixed colour scheme. Each photograph was displayed for 5 s, and the position and duration of each fixation were recorded. Although there has been interest in the analysis of 'task-based' ocular movements such as searching a photograph for a point in interest, in this experiment no instructions were given leading to a 'free-viewing' scenario. For each individual fixation, a data entry includes the horizontal and vertical position of the fixation, the duration in milliseconds, the fixation's sequence number, the subject identifier, the colour scheme, the image number and the orientation of the image. A more detailed explanation of the data, the experiment and the method of collection can be found in Ho-Phuoc *et al.* (2012).

In this paper, Section 2 introduces a discrete time Markov model for the observed sequences of ocular fixations. The states are determined through k-means clustering where cross-validation is used to determine the optimal number of clusters. A further investigation of alternative clustering methods, a closer look at the temporal dependence of the point process, a *post hoc* look at the Markov transition probabilities and a display of the best and worst scoring photographs under our model can be found in Sections 2.1, 2.2, 2.3 and 2.4 respectively. Section 3 proposes reworking the discrete time model as a continuous time Markov chain through a closer analysis of the fixations' durations. Lastly, Section 4 concludes with potential applications.

The images that are analysed in the paper can be obtained from `https://github.com/cachelack/eyeFixationData` and the programs that were used to analyse them from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Discrete time Markov model

Consider a sequence of n fixation positions from a single subject, $X_1, \ldots, X_n \in \mathbb{R}^2$, as a point process in $\mathbb{R}^2$ and an associated sequence of n states $S_1, \ldots, S_n \in \{1, \ldots, k\}$. We shall model this state sequence as a Markov chain jumping between k different clusters corresponding to interesting parts of the photograph. The fixation sequence will then be random observations conditioned on the current state of the Markov chain. The model selection will decide between such models for $k = 1, \ldots, 10$. The case $k = 1$, our null model with which to compare the others, is the naive model that the X_t for $t = 1, \ldots, n$ are independent and identically distributed draws from some underlying density $f(x)$. For $k \geqslant 2$, we suppose a finite mixture model with k constituent densities $f_1, \ldots, f_k$ corresponding to which part of the image the eye is focusing on. In this model, the states evolve via a Markov chain with X_t for $t = 1, \ldots, n$ given by an independent random draw from $f_{S_t}(x)$.

These constituent densities were modelled empirically by clustering the fixation locations from nine of the 10 subjects; the model was then tested on the 10th. Cross-validation was performed across all training subjects to optimize this model. Let $X_1, \ldots, X_n$ be the test sequence of fixation locations for a single subject with n being the total number of fixations made by this subject, and let $Y_t^{(j)}$ be fixation t of subject j from the training set where $t = 1, \ldots, n_j$ with n_j being the total number of fixations made by subject j. A Bayes factor was computed for each subject, and the results were averaged into a final score for each picture. The training fixation points were clustered via k-means clustering with 10 random starts. Other clustering methods are discussed in Section 2.1. For each cluster, a two-dimensional kernel density estimate with Gaussian kernel was computed. An example of these clusters and density estimates can be seen in Fig. 2. Each fixation in the test set was assigned a cluster on the basis of proximity to the cluster centre. A k-nearest-neighbours classifier was also implemented to assign clusters but returned very similar results.

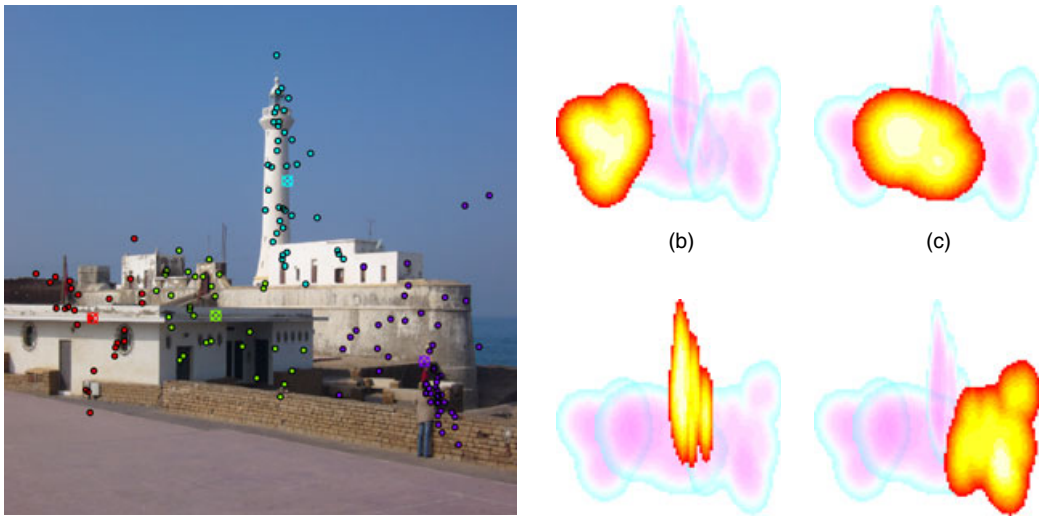


Fig. 2. (a) The four clusters of fixation locations for a given image and (b)–(e) the corresponding kernel density estimates for the fixation location model: fixations on image 25, BF = 0.00267

The observed initial states and transitions between states were treated as observations from a multinomial random variable with a Dirichlet conjugate prior. Specifically, the Markov initial state probabilities, π_i for $i \in \{1,\ldots,k\}$, and Markov transition matrix, $p_{i,i'}$ for i, $i' \in \{1,\ldots,k\}$, were treated as Dirichlet random variables with the Jeffreys prior and updated by the nine subjects in the training data. Let c_i be the number of initial fixations $Y_1^{(j)}$ in state i, and let $m_{i,i'}$ be the number of observed transitions from $Y_{t-1}^{(j)} \in S_i$ to $Y_t^{(j)} \in S_{i'}$ for $t = 2,\ldots,n_j$. The posteriors are

$$\begin{aligned} \pi &\sim \text{Dirichlet}(0.5+c_1,\ldots,0.5+c_k), \\ p_{i,\cdot} &\sim \text{Dirichlet}(0.5+m_{i,1},\ldots,0.5+m_{i,k}). \end{aligned}$$

Therefore, the Bayes factor is

$$\text{BF} = \frac{P(X_t|Y_t,k=1)}{P(X_t|Y_t,k)} = \frac{\prod_{t=1}^{n} f(X_t)}{E_{\pi,p}\left\{\pi_{s_1} f_{s_1}(X_1) \prod_{t=2}^{n} p_{s_{t-1},s_t} f_{s_t}(X_t)\right\}}$$

where the expectation is taken with respect to the Dirichlet posterior. In practice, this value is approximated via Monte Carlo integration.

To identify a difference in the computed Bayes factors between each of the three colour schemes, a three-category analysis of variance was run on the $\log_2(\text{BF})$s, yielding a strongly significant p-value of 1.87×10^{-5}. Thus, the null hypothesis that the three sets of $\log_2(\text{BF})$s have equal means is rejected. Furthermore, a *post hoc* Tukey test was run to construct three pairwise confidence intervals for the differences of the means with a 95% familywise coverage probability. This results in the following simultaneous confidence intervals and p-values:

normal–abnormal,	[−2.01, −1.55]	p-value 0.95;
normal–greyscale,	[−4.79, −1.24]	p-value 0.000 27;
abnormal–greyscale,	[−5.02, −1.47]	p-value 0.000 079.

Consequently, the presence of colour, whether normal or not, results in the majority of images scoring a relatively smaller Bayes factor than in the greyscale setting. This implies that our proposed Markov model better describes the ocular fixation data when colour information is present.

Furthermore, the Bayes factors for colour and greyscale images separate sufficiently well that this model applied to observed sequences of fixations can be used as a weak classifier for whether or not the subjects are observing an image with colour information. Indeed, over all of the 60 pictures and three colour schemes, 14 normal, 13 abnormal, but only one greyscale picture scored a Bayes factor less than 0.01. The threshold that most separates this data set is 0.2, which correctly separates 66% of the normal and 71% of the abnormal schemes from the greyscale images. Thresholding the Bayes factor as a classification criterion for whether or not the observed photograph has colour, is normal or abnormal results in the receiver operating characteristic curves of Fig. 3. The collection of displayed receiver operating characteristic curves includes two implementing k-means clustering where inclusion is based on either proximity to the cluster centre or the k-nearest-neighbours method. Two hierarchical clustering methods are also included, which will be discussed more in Section 2.1. Here, 'true positive' refers to the percentage of coloured photographs with Bayes factor below the threshold and 'false positive' for the percentage of greyscale photographs below the threshold.

Ultimately, the ocular fixation data for greyscale images do not provide evidence that the Markov model is a better explanation than merely modelling the fixations as a collection of independent random draws. Although it is doubtful that the removal of colour actually reduces the eye to a pure random search, this result does support the drop in efficiency that is witnessed in the greyscale setting. In contrast, the coloured cases are well modelled as if jumps between interesting regions of the image occur in a Markovian fashion. This suggests that the absence of colour can make it more difficult for subjects to identify and scan through interesting parts of an image.

2.1. Clustering methods

The use of k-means clustering with Euclidean distance puts a heavy assumption on our model. Specifically, this approach partitions a photograph into Voronoi cells, which are by design all convex polygons. This approach strives to construct spherical and similarly sized clusters specifically removing the possibility of non-convex or nested clusters. In light of this, a variety of agglomerative hierarchical clustering methods were also tested.

These 'bottom-up' hierarchical clustering methods begin with each fixation occupying its own cluster. The methods iteratively combine clusters on the basis of a combining criterion and an underlying metric. In our analysis, the metrics chosen to test were the Manhattan or L^1-, the Euclidean or L^2- and the maximum or L^∞-distances. The linkage methods chosen were Ward's minimum variance method (Ward, 1963; Murtagh and Legendre, 2014), complete linkage clustering and the unweighted pair group method with arithmetic mean (Sokal and Michener, 1958). See section 14.3 of Hastie *et al*. (2005) or section 8.5 of Legendre and Legendre (2012) for an overview of such methods.

Of the various combinations of such metrics and linkage criteria, none performed noticeably better than k-means, and many combinations performed worse. Fig. 3 includes two receiver operating characteristic curves by using Ward's method with L^∞- and L^1-distances. As the clusters formed via hierarchical clustering need not be convex, the k-nearest-neighbours method was used to determine to which cluster a given fixation from the testing set belonged. A variety

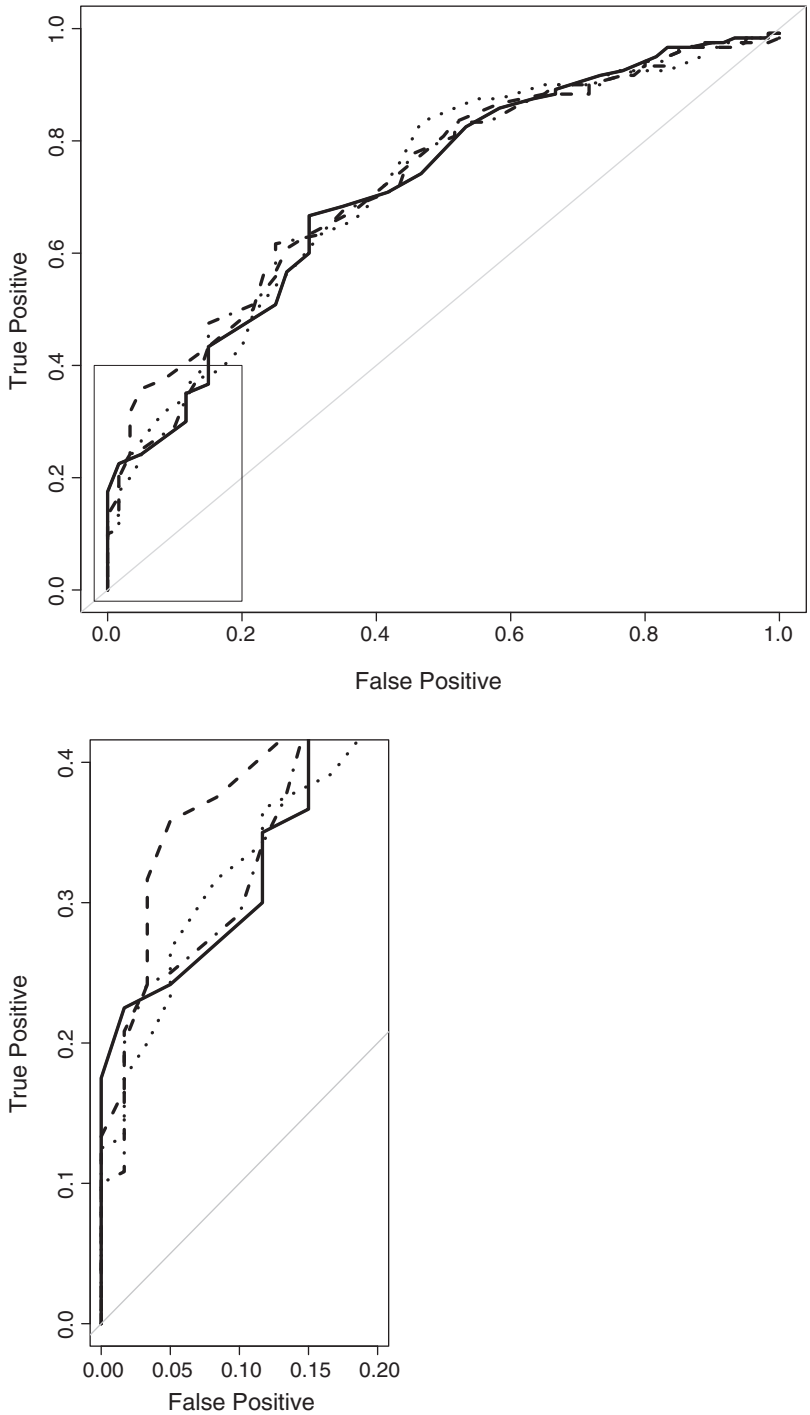


Fig. 3. Receiver operating characteristic curves for thresholding on the Bayes factor to determine whether the photograph being viewed is in colour (true positive) or greyscale (false positive): four clustering methods are plotted with similar results (———, *k*-means, cluster centres; - - - -, *k*-means, *k*-nearest-neighbour method; · · · · · ·, L^{∞}-distance, Ward (1963); · — · —, L^{1}-distance, Ward (1963))

of values k were tested. Lastly, a Gaussian mixture model was also fitted to the data via an EM algorithm. The performance of our model under this clustering method was also similar to the k-means results.

2.2. Temporal dependence

The Markov model proposed confirms strongly the spatial dependence of the sequences of fixations, but it can further be used to identify temporal dependence in such sequences. A threshold to delineate significant Bayes factors was chosen as 0.15. Of the 180 data sets, which correspond to 60 pictures and three colour schemes, 93 surpassed this threshold, providing moderate to strong evidence in favour of the Markov model. For each of those 93 images, the corresponding fixations were partitioned into two groups: those that occurred in the first 2 s of the experiment, and those that occurred after the 2-s mark. Given a model with k states, counts of the observed states from the two disjoint timeframes were compared in a $2 \times k$ contingency table by using the Kruskal–Wallis criterion to test for an association (Kruskal and Wallis, 1952). Simulations from the exact null distribution of the Kruskal–Wallis test statistic were performed by the R package `kSamples` (Scholz and Zhu, 2016) and were used to compute p-values for each of the 93 tables. The Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) was employed to control the false discovery rate. Of the 93 p-values, 44 were deemed significant with a false discovery rate of 0.05. This provides strong evidence that in many situations there is a temporal dependence regarding in which region the eye will fixate.

However, although the subjects may begin by fixating in one specific cluster only to move to another for the latter half of the time interval, there is no significant evidence to claim that the Markov transition probabilities evolve over time. Indeed, the above procedure was run again but on the matrix of observed transition counts for each of the 93 data sets under scrutiny. The Kruskal–Wallis test was now run on each row of the matrix as a separate block and those statistics were combined to obtain an overall test statistic for each image. After applying the Benjamini–Hochberg test with a false discovery rate of 0.05 again, none of the p-values were sufficiently small to be designated as significant. Thus, although the human eye is shown to have a preference for in which cluster to start and in which cluster to end, transition probabilities remain fixed over the timespan of this experiment.

As a concrete example, the above procedure applied to the fixations on image 34 under the abnormal colour scheme resulted in this image being partitioned into three clusters. Counting the members of each cluster and the transitions between each pair of clusters resulted in Table 1. During the first 2 s, more of the fixations occurred in cluster 3 than in cluster 1. During the subsequent time interval, the reverse behaviour was witnessed. The contingency-table-based comparison of the cluster counts returned an extreme p-value of 9.9×10^{-12} whereas the contingency table test between the transition counts returned a much less exciting p-value of 0.16. Note that, in Table 1 in the time $\geqslant 2$ s section, the sum of the 'total counts' is precisely 10 more than the sum of the 'transitions' as each of the 10 subjects had a final fixation with no subsequent transition.

2.3. Analysis of transitions

The saccades, i.e. the rapid eye movement from one fixation to the next, have been intensely studied in their own right. Brockmann and Geisel (2000) and Boccignone and Ferraro (2004) treated such saccadic eye movements as a stochastic jump process working under the assumption that the eye has evolved a search strategy to minimize the required time to absorb the necessary information in the visual scene confronting it.

Table 1. Counts of the observed fixations and transitions of the 10 subjects on image 34 with abnormal colours partitioned by two time intervals†

Cluster	*Results for time* $< 2s$				*Results for time* $\geqslant 2s$			
	Total counts	*Transitions*			*Total counts*	*Transitions*		
		1	*2*	*3*		*1*	*2*	*3*
1	5	3	0	2	44	33	6	3
2	24	3	17	4	28	7	15	2
3	45	1	11	33	14	1	2	7

†The 'total counts' column refers to the total observed fixations in each cluster during the specific time interval. The remaining columns refer to the number of observed transitions from the row cluster to the column cluster.

For images that strongly fit the above Markov model, the saccades follow a natural mixture model. For example, if the Markov point process is supported on two states, then the eye can choose to stay in the same state (i.e. a short saccade) or transition to the other state (i.e. a long saccade). This observation is in line with the work of Tatler and Vincent (2008) who referred to 'distinct modes' of local search and global search strategies chosen by the human eye when viewing complex scenes. Boccignone and Ferraro (2013) compared the eye's search strategy with the 'feed-and-fly' model of animal foraging referring similarly to local searches and large jumps. Such dichotomous behaviour is readily evident in Fig. 4, which displays the two clusters of fixations for image 9 in Fig. 4(a) and a kernel density estimate for the distribution of the saccade lengths in Fig. 4(b). The kernel density estimate was computed with a Gaussian kernel using the Sheather–Jones method of bandwidth selection (Sheather and Jones, 1991).

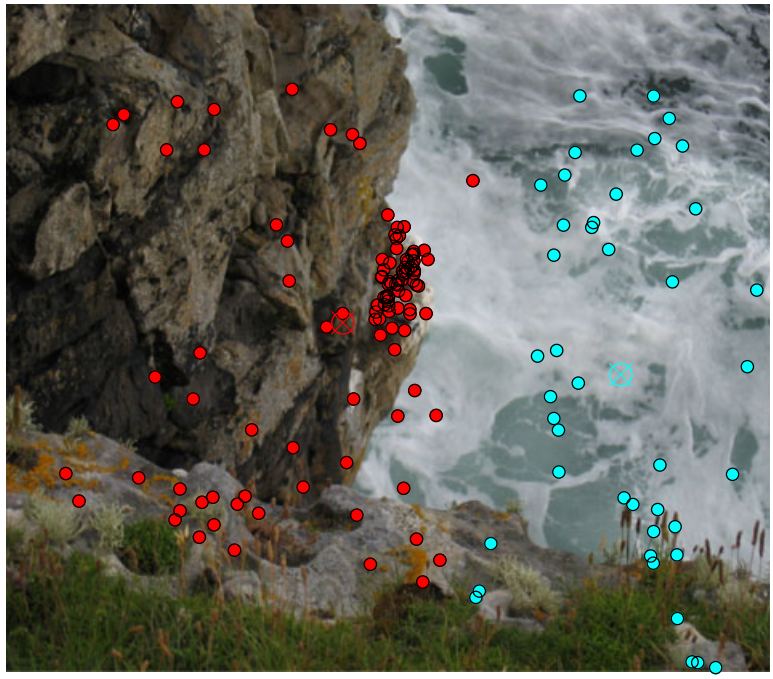
For a *post hoc* look at the observed transitions between states, we shall consider image 25 under the normal colour scheme, which is displayed in Fig. 2. Running the above analysis yielded a decisively strong Bayes factor of 0.0009 in favour of the model $k = 4$ over the model $k = 1$. The states are coloured red, green, cyan and purple moving from left to right across the image.

For each of the 10 subjects, the maximum likelihood estimate of the initial probabilities and transition matrix from the Dirichlet posteriors were averaged into

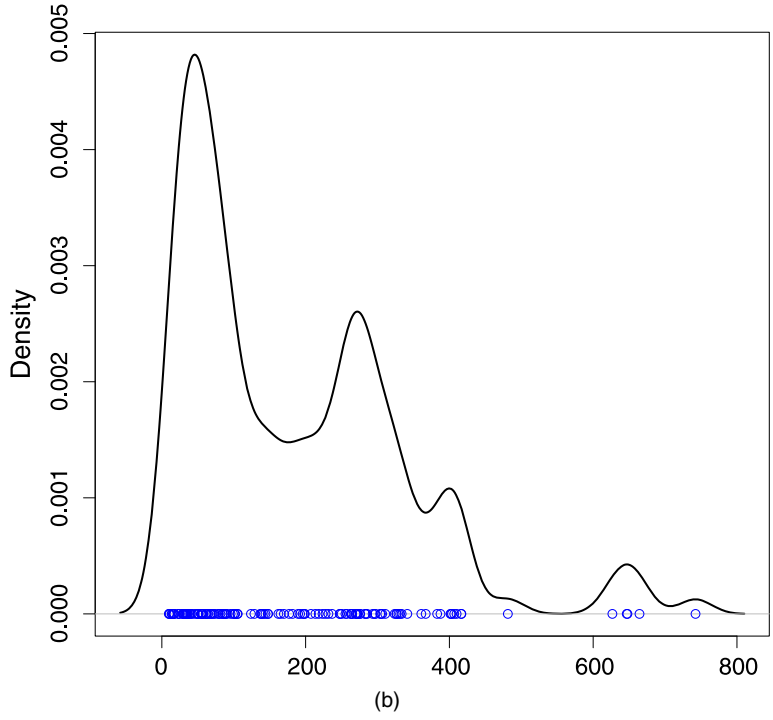
$$\pi = \begin{pmatrix} 0.05 \\ 0.45 \\ 0.13 \\ 0.37 \end{pmatrix},$$

$$p = \begin{pmatrix} 0.51 & 0.29 & 0.14 & 0.06 \\ 0.30 & 0.26 & 0.24 & 0.20 \\ 0.07 & 0.18 & 0.58 & 0.18 \\ 0.05 & 0.11 & 0.13 & 0.70 \end{pmatrix}.$$

Here, states 1, 3 and 4 fall into the often seen pattern of having probability higher than 50% of remaining in the same state and of having other transition probabilities that roughly decrease as the distance between clusters increases. This behaviour indicates that the human eye will spend a short sequence of fixations examining one specific region of the photograph before jumping to another and subsequently remaining there for another short sequence of fixations.

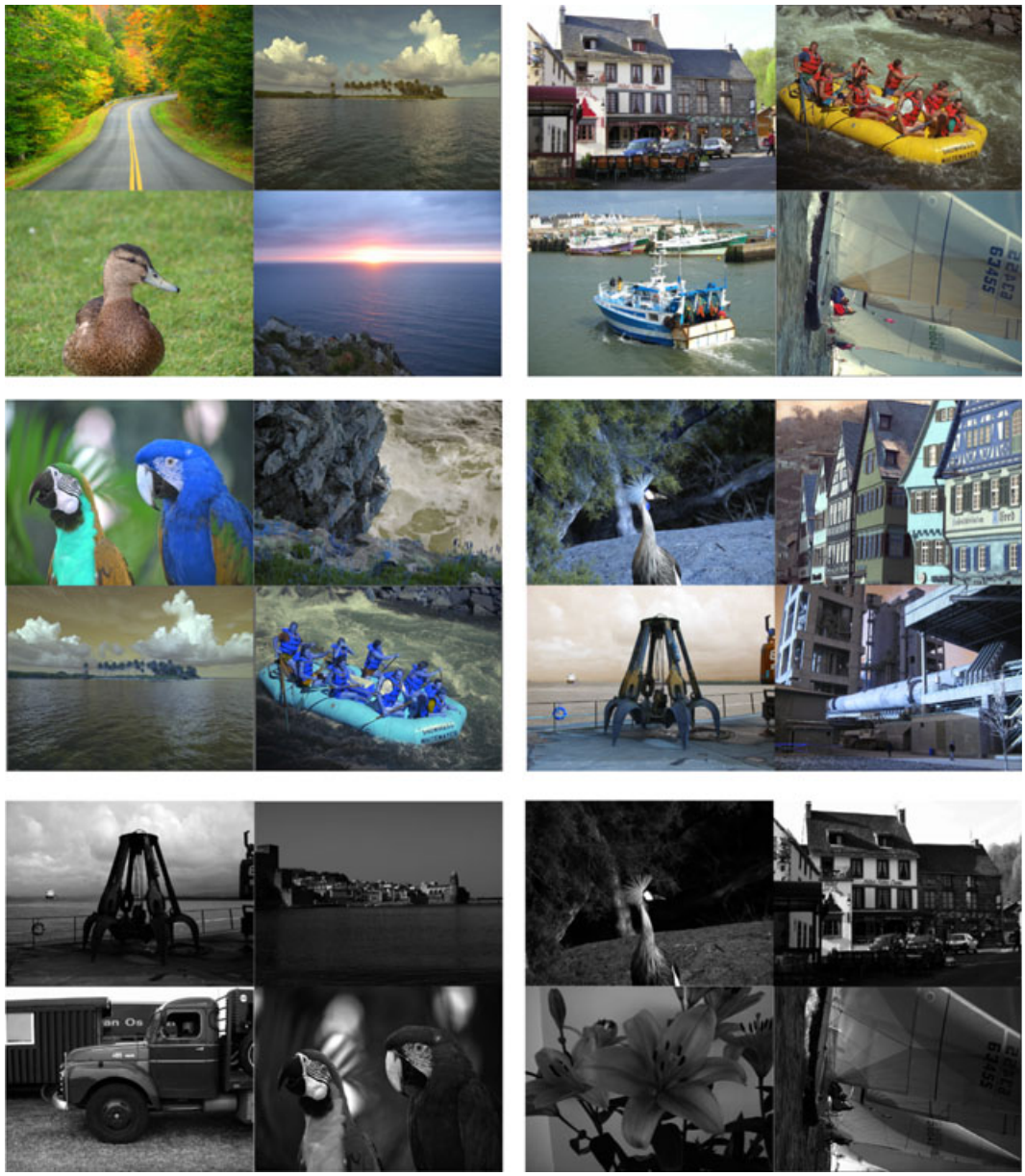


(a)



(b)

Fig. 4. (a) Two clusters of fixations, image 9, and (b) a kernel density plot of the saccade lengths, image 9 ($N = 141$; bandwidth 22.63)



(a) (b)

Fig. 5. (a) The four worst scoring pictures and (b) the four best scoring pictures for the three colour schemes

2.4. The best and the worst

Using the strongest Bayes factor for each photograph, we can rank them in order of which best fits the Markov model for $k \geqslant 2$ clusters. The four best and worst photographs for each colour scheme are depicted in Fig. 5. The normal coloured images are reasonably partitioned as the worst scoring images contain a singular point of focus whereas the best scorers contain multiple objects on which to fixate such as text and people. The abnormal setting gives similar results barring the photograph of eight people in a raft, which scores strongly under normal colours but produced the single worst score under the abnormal scheme. Presumably, the abnormal colour scheme most disorients the brain when it is applied to objects with a narrow range of expected colours such as human faces, which are generally not blue.

3. Continuous time Markov model

The most blatant omission from the previously described model is the time that is spent at each fixation. In fact, evidence suggests that the colour scheme has a significant effect on the fixation durations. In the initial investigation of this data set by Ho-Phuoc *et al*. (2012), Kolmogorov–Smirnov tests were run between each pair of empirical distributions for the overall fixations durations. They reported no significant difference between the normal and abnormal settings, but they reported high significance between the greyscale and each of the coloured settings. We shall now consider the fixations as a continuous time point process. It will be demonstrated that this process is temporally homogeneous in the sense that the jumps occur at a fixed exponential rate in accordance with earlier research (Salthouse and Ellis, 1980; Harris *et al*., 1988; Manor and Gordon, 2003).

For $t \in [0, 5]$, let $X_t \in \mathbb{R}^2$ be a continuous time point process corresponding to a subject's fixation location at any time t. If X_t achieves $n + 1$ unique values in the time interval, then denote the jump times between adjacent values as $T_1, \ldots, T_n$ where $0 < T_1 < \ldots < T_n < 5$. For all 180 data sets, the Anderson–Darling test was used to compare the empirical distribution of these random jump times with a uniform distribution. To avoid edge effects, the time interval of scrutiny was truncated to jumps occurring in between 0.5 and 4.0 s before the test was run. Even without considering multiple-testing corrections, the smallest p-value of the 180 was an unexciting 0.067. Each of the three sets of 60 p-values corresponding to one of the colour schemes was aggregated by using Fisher's method for combining independent tests to check for consistent but weak deviations from uniformity. However, the resulting aggregated p-values for normal, abnormal and greyscale images were the almost too flat 0.63, 0.993 and 0.998 respectively. This provides strong evidence that, albeit after removing the edge effects, the fixations are uniformly distributed in time and thus occur at a constant exponential rate.

The jump rate was estimated in the usual way by counting the total number of jumps that occurred in the time interval [0.5, 4.0]. This was performed for each of the 180 images and averaged over each of the three colour scheme sets. The resulting rates are 3.65, 3.62 and 3.25 fixations per second for normal, abnormal and greyscale colours respectively. Similarly to the analysis of the Bayes factors in Section 2, an analysis of variance was run to compare the three sets of 60 fixations rates. This resulted in a decisively strong p-value of 4.6×10^{-11}. Computing the *post hoc* 95% Tukey confidence intervals gives the following results:

$$\begin{aligned}
&\text{normal–abnormal,} && [-0.12, -0.17] && p\text{-value } 0.91;\\
&\text{normal–greyscale,} && [-0.54, -0.26] && p\text{-value less than } 10^{-7};\\
&\text{abnormal–greyscale,} && [-0.51, -0.23] && p\text{-value less than } 10^{-7}.
\end{aligned}$$

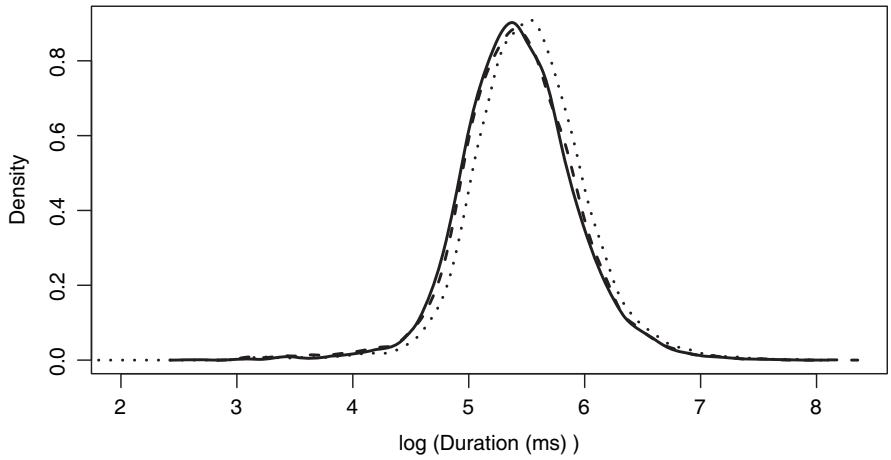


Fig. 6. Kernel density estimates of the distributions of the logarithm of the fixation durations separated by colour scheme: ———, normal; - - - -, abnormal; · · · · · ·, greyscale

This demonstrates that there is a significant drop in the fixation rate in the greyscale setting compared with when colour information is present. Equivalently, in Fig. 6, the density plot of the greyscale durations is noticeably shifted to the right of the other two density curves.

4. Summary and discussion

Ocular fixation data yield both a challenging and a useful analysis. We demonstrated that, in the presence of colour information whether normal or abnormal, a sequence of fixations from a human eye can be modelled as a Markov point process and that the fixations occur at a constant rate. Furthermore, this Markov model breaks down when such colour information is removed from the image, and the fixation rate drops significantly. Given that we believe this model, a thorough analysis of the Dirichlet posteriors could yield interesting insight into how different photographs are treated by the human eye.

This model has the potential to lead to future applications such as a passive diagnostic test for sudden loss of colour vision, i.e., even with the 60 miscellaneous photographs of the given data set, it is still possible to classify which observers are looking at colour images and which are not. Indeed, simultaneously thresholding on the Bayes factors from Section 2 and on the fixation rates from Section 3 allows for a true empirical classification rate of 0.317 for a set false positive rate of 0.05. With a carefully constructed set of colour photographs to elicit Markovian eye movements, one could then use data collected from healthy eyes to train a classifier to determine whether a patient can see colour or not with no other active participation from the patient besides staring at the set of diagnostic photographs.

Ultimately, there is still room for further analysis in at least two areas of note. In our model, the spatial and temporal aspects of the process are treated separately. A more sophisticated point process model could take such interdependences between these two dimensions into account. Secondly, attempting to construct such a model from image information rather than fixation data could offer insight into saliency maps and lead to more complex and interesting Markov states than the convex polygons from the k-means approach.

Acknowledgements

The authors acknowledge the Young Statisticians Section and Research Section of the Royal Statistical Society and the contest sponsor, Select Statistics, for constructing a very enjoyable and intellectually exciting statistical analytics challenge. We also thank the referees for the invaluable comments that they provided.

This work was supported by UK Engineering and Physical Sciences Research Council grant EP/H023348/1 for the University of Cambridge Centre for Doctoral Training, Cambridge Centre for Analysis.

References

Achanta, R., Estrada, F., Wils, P. and Süsstrunk, S. (2008) Salient region detection and segmentation. In *Proc. Int. Conf. Computer Vision Systems* (eds A. Gasteratos, M. Vincze and J. K. Tsotsos), pp. 66–75. New York: Springer.

Amano, K. and Foster, D. H. (2014) Influence of local scene color on fixation position in visual search. *J. Opt. Soc. Am.* A, **31**, A254–A262.

Avidan, S. and Shamir, A. (2007) Seam carving for content-aware image resizing. *ACM Trans. Graph.*, **26**, no. 3, article 10.

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

Baddeley, R. J. and Tatler, B. W. (2006) High frequency edges (but not contrast) predict where we fixate: a bayesian system identification analysis. *Visn Res.*, **46**, 2824–2833.

Barthelmé, S., Trukenbrod, H., Engbert, R. and Wichmann, F. (2013) Modeling fixation locations using spatial point processes. *J. Visn*, **13**, article 1.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc.* B, **57**, 289–300.

Berthelsen, K. K. and Møller, J. (2008) Non-parametric bayesian inference for inhomogeneous Markov point processes. *Aust. New Zeal. J. Statist.*, **50**, 257–272.

Boccignone, G. and Ferraro, M. (2004) Modelling gaze shift as a constrained random walk. *Physica* A, **331**, 207–218.

Boccignone, G. and Ferraro, M. (2013) Feed and fly control of visual scanpaths for foveation image processing. *Ann. Telecommuns*, **68**, 201–217.

Brix, A. and Diggle, P. J. (2001) Spatiotemporal prediction for log-Gaussian Cox processes. *J. R. Statist. Soc.* B, **63**, 823–841.

Brockmann, D. and Geisel, T. (2000) The ecology of gaze shifts. *Neurocomputing*, **32**, 643–650.

Bruce, N. D. and Tsotsos, J. K. (2009) Saliency, attention, and visual search: an information theoretic approach. *J. Visn*, **9**, article 5.

Cerf, M., Frady, E. P. and Koch, C. (2009) Faces and text attract gaze independent of the task: experimental data and computer model. *J. Visn*, **9**, article 10.

Chen, L.-Q., Xie, X., Fan, X., Ma, W.-Y., Zhang, H.-J. and Zhou, H.-Q. (2003) A visual attention model for adapting images on small displays. *Multmed. Syst.*, **9**, 353–364.

Cox, D. R. and Isham, V. (1980) *Point Processes*. Boca Raton: Chapman and Hall–CRC.

Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *A. Rev. Neursci.*, **18**, 193–222.

Diggle, P. J. (2003) *Statistical Analysis of Spatial Point Patterns*. New York: Academic Press.

Diggle, P. J., Moraga, P., Rowlingson, B. and Taylor, B. M. (2013) Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. *Statist. Sci.*, **28**, 542–563.

Dominy, N. J. and Lucas, P. W. (2001) Ecological importance of trichromatic vision to primates. *Nature*, **410**, 363–366.

Frey, H.-P., Honey, C. and König, P. (2008) What's color got to do with it?: the influence of color on visual attention in different categories. *J. Visn*, **8**, article 6.

Good, I. J. (1967) A Bayesian significance test for multinomial distributions (with discussion). *J. R. Statist. Soc.* B, **29**, 399–431.

Hacisalihzade, S. S., Stark, L. W. and Allen, J. S. (1992) Visual perception and sequences of eye movement fixations: a stochastic modeling approach. *IEEE Trans. Syst. Man Cyber.*, **22**, 474–481.

Hamel, S., Guyader, N., Pellerin, D. and Houzet, D. (2014) Color information in a model of saliency. In *Proc. 22nd Eur. Conf. Signal Processing, Lisbon*, pp. 226–230. New York: Institute of Electrical and Electronics Engineers.

Harris, C. M., Hainline, L., Abramov, I., Lemerise, E. and Camenzuli, C. (1988) The distribution of fixation durations in infants and naive adults. *Visn Res.*, **28**, 419–432.

Hastie, T., Tibshirani, R., Friedman, J. and Franklin, J. (2005) The elements of statistical learning: data mining, inference and prediction. *Math. Intell.*, **27**, 83–85.

Ho-Phuoc, T., Guyader, N., Landragin, F. and Guérin-Dugué, A. (2012) When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *J. Visn*, **12**, article 4.
Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008) *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: Wiley.
Itti, L. and Koch, C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Visn Res.*, **40**, 1489–1506.
Jensen, E. B. V. and Nielsen, L. S. (2000) Inhomogeneous Markov point processes by transformation. *Bernoulli*, **6**, 761–782.
Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *J. Am. Statist. Ass.*, **90**, 773–795.
Ko, B. C. and Nam, J.-Y. (2006) Object-of-interest image segmentation based on human attention and semantic region clustering. *J. Opt. Soc. Am.* A, **23**, 2462–2470.
Koch, C. and Ullman, S. (1987) Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of Intelligence* (ed. L. M. Vaina), pp. 115–141. Berlin: Springer.
Kruskal, W. H. and Wallis, W. A. (1952) Use of ranks in one-criterion variance analysis. *J. Am. Statist. Ass.*, **47**, 583–621.
Kümmerer, M., Theis, L. and Bethge, M. (2014) Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet. *Preprint arXiv:1411.1045*. Werner Reichardt Centre for Integrative Neuroscience, University of Tubingen, Tubingen.
Legendre, P. and Legendre, L. F. (2012) *Numerical Ecology*. Amsterdam: Elsevier.
Le Meur, O. and Coutrot, A. (2016) Introducing context-dependent and spatially-variant viewing biases in saccadic models. *Visn Res.*, **121**, 72–84.
Manor, B. R. and Gordon, E. (2003) Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *J. Neursci. Meth.*, **128**, 85–93.
McLachlan, G. and Peel, D. (2004) *Finite Mixture Models*. New York: Wiley.
Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998) Log Gaussian Cox processes. *Scand. J. Statist.*, **25**, 451–482.
Murtagh, F. and Legendre, P. (2014) Wards hierarchical agglomerative clustering method: which algorithms implement Wards criterion? *J. Classificn*, **31**, 274–295.
Salthouse, T. A. and Ellis, C. L. (1980) Determinants of eye-fixation duration. *Am. J. Psychol.*, **93**, 207–234.
Scholz, F. and Zhu, A. (2016) kSamples: K-sample rank tests and their combinations. *R Package Version 1.2-4*. (Available from `https://CRAN.R-project.org/package=kSamples`.)
Schütz, A. C., Braun, D. I. and Gegenfurtner, K. R. (2011) Eye movements and perception: a selective review. *J. Visn*, **11**, article 9.
Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Statist. Soc.* B, **53**, 683–690.
Sokal, R. R. and Michener, C. D. (1958) A statistical method for evaluating systematic relationships. *Univ. Kansas Scient. Bull.*, **28**, 1409–1438.
Stark, L. W. and Ellis, S. R. (1981) Scanpaths revisited: cognitive models direct active looking. In *Eye Movements, Cognition and Visual Perception* (ed. D. F. Fisher), pp. 193–226. Mahwah: Erlbaum.
Sumner, P. and Mollon, J. (2000) Catarrhine photopigments are optimized for detecting targets against a foliage background. *J. Exptl Biol.*, **203**, 1963–1986.
Tatler, B. W., Hayhoe, M. M., Land, M. F. and Ballard, D. H. (2011) Eye guidance in natural vision: reinterpreting salience. *J. Visn*, **11**, article 5.
Tatler, B. W. and Vincent, B. T. (2008) Systematic tendencies in scene viewing. *J. Eye Movmnt Res.*, **2**, no. 2.
Taylor, B. M., Davies, T. M., Rowlingson, B. S. and Diggle, P. J. (2013) lgcp: an R package for inference with spatial and spatio-temporal log-Gaussian Cox processes. *J. Statist. Softwr.*, **52**, 1–40.
Wang, Z., Lu, L. and Bovik, A. C. (2003) Foveation scalable video coding with automatic fixation selection. *IEEE Trans. Im. Process.*, **12**, 243–254.
Ward, J. H. (1963) Hierarchical grouping to optimize an objective function. *J. Am. Statist. Ass.*, **58**, 236–244.
Zuber, B. L. (1981) *Models of Oculomotor Behavior and Control*. Boca Raton: CRC Press.