

Spatial point processes for eye-movement data

Simon Barthelmé, University of Geneva

July 3, 2012

The plan

- ▶ I am working on a R package called *mpp* that will provide tools for Bayesian inference for certain kinds of point process models.

The plan

- ▶ I am working on a R package called *mpp* that will provide tools for Bayesian inference for certain kinds of point process models.
- ▶ These models are actually very general but because I promised eye movement data that's what you'll get

The plan

- ▶ I am working on a R package called *mpp* that will provide tools for Bayesian inference for certain kinds of point process models.
- ▶ These models are actually very general but because I promised eye movement data that's what you'll get
- ▶ It is hard to introduce both the high-level concepts and their implementation, so I'll focus on why these models are interesting.

The plan

- ▶ I am working on a R package called *mpp* that will provide tools for Bayesian inference for certain kinds of point process models.
- ▶ These models are actually very general but because I promised eye movement data that's what you'll get
- ▶ It is hard to introduce both the high-level concepts and their implementation, so I'll focus on why these models are interesting.
- ▶ I'll still try to give some idea of *mpp*'s interface.

The saliency map hypothesis

- ▶ Main theoretical idea in the bottom-up camp: the visual system determines where to look according to a saliency map that is mainly a function of local image properties (Koch & Ullmann, 1985).

The saliency map hypothesis

- ▶ Main theoretical idea in the bottom-up camp: the visual system determines where to look according to a saliency map that is mainly a function of local image properties (Koch & Ullmann, 1985).



Visual Scene

The saliency map hypothesis

- ▶ Main theoretical idea in the bottom-up camp: the visual system determines where to look according to a saliency map that is mainly a function of local image properties (Koch & Ullmann, 1985).



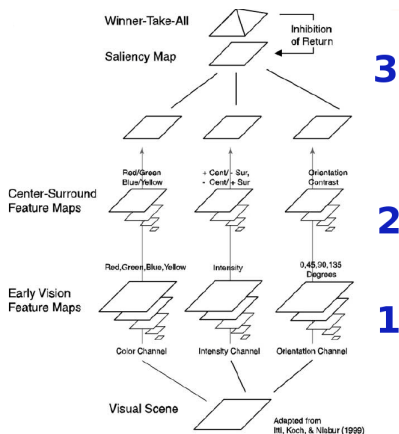
Visual Scene



Saliency Map

Saliency models

- ▶ Itti & Koch model (here): extract simple local properties of the image (1), see where they change (2), combine (3).



Objective

- ▶ Models of low-level saliency are *extremely* popular
 - ▶ Google scholar citation count for the three most-cited papers by Itti&Koch $\sim 10,000$ (Google scholar)

Objective

- ▶ Models of low-level saliency are *extremely* popular
 - ▶ Google scholar citation count for the three most-cited papers by Itti&Koch ~ 10,000 (Google scholar)
 - ▶ Very strong demand from engineering (compression, steganography, computer vision) and marketing.

Objective

- ▶ Models of low-level saliency are *extremely* popular
 - ▶ Google scholar citation count for the three most-cited papers by Itti&Koch $\sim 10,000$ (Google scholar)
 - ▶ Very strong demand from engineering (compression, steganography, computer vision) and marketing.
- ▶ How well do these models really predict the data? Can we improve on them?

Objective

- ▶ Models of low-level saliency are *extremely* popular
 - ▶ Google scholar citation count for the three most-cited papers by Itti&Koch $\sim 10,000$ (Google scholar)
 - ▶ Very strong demand from engineering (compression, steganography, computer vision) and marketing.
- ▶ How well do these models really predict the data? Can we improve on them?
- ▶ We need a proper statistical framework.

The data

- ▶ “Free-viewing” experiments: sit people down in front of a computer screen, show them natural images, record their eye movements. No task, just “explore the image”.

The data

- ▶ “Free-viewing” experiments: sit people down in front of a computer screen, show them natural images, record their eye movements. No task, just “explore the image”.
- ▶ We use data from Kienzle et al. (2009).

The data

- ▶ “Free-viewing” experiments: sit people down in front of a computer screen, show them natural images, record their eye movements. No task, just “explore the image”.
- ▶ We use data from Kienzle et al. (2009).

The data

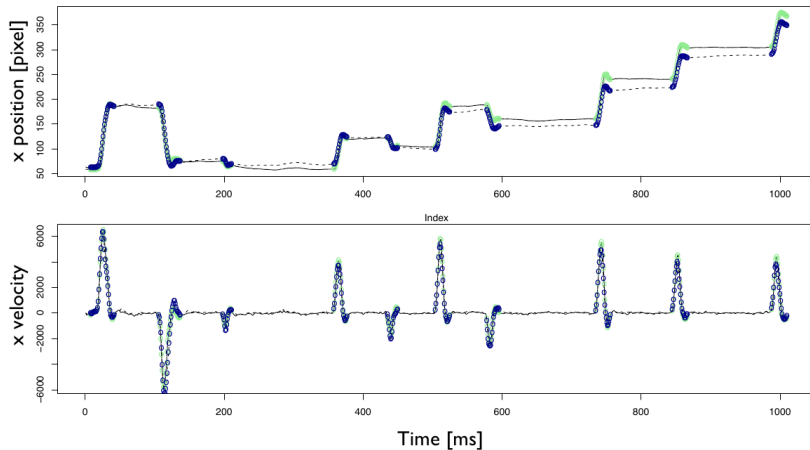
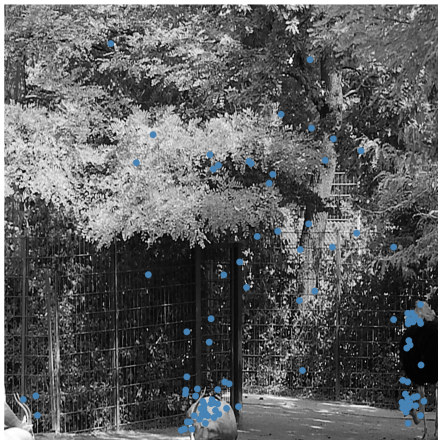


Figure by Jochen Laubrock

The data



Spatial point processes

- ▶ Point processes are the core tool of *spatial statistics* (Illian *et al.*, 2008).

Spatial point processes

- ▶ Point processes are the core tool of *spatial statistics* (Illian *et al.*, 2008).
- ▶ They are now used quite often in other branches of biology (ecology), but have yet to be applied to neuroscience and psychology.

Spatial point processes

- ▶ Point processes are the core tool of *spatial statistics* (Illian *et al.*, 2008).
- ▶ They are now used quite often in other branches of biology (ecology), but have yet to be applied to neuroscience and psychology.
- ▶ I'll try to argue that they provide a very good way to think about some aspects of eye movement data, esp. the analysis of fixation locations.

Why are point processes useful in that context?

- ▶ Take the idea that “my model of saliency maps will help me predict where people fixate”

Why are point processes useful in that context?

- ▶ Take the idea that “my model of saliency maps will help me predict where people fixate”
 - ▶ How exactly?

Why are point processes useful in that context?

- ▶ Take the idea that “my model of saliency maps will help me predict where people fixate”
 - ▶ How exactly?
 - ▶ We need a quantitative version of that statement

Why are point processes useful in that context?

- ▶ Take the idea that “my model of saliency maps will help me predict where people fixate”
 - ▶ How exactly?
 - ▶ We need a quantitative version of that statement
 - ▶ Plenty of confusion in the literature as to how we make this quantitative, exactly.

Why are point processes useful in that context?

- ▶ Take the idea that “my model of saliency maps will help me predict where people fixate”
 - ▶ How exactly?
 - ▶ We need a quantitative version of that statement
 - ▶ Plenty of confusion in the literature as to how we make this quantitative, exactly.
- ▶ I'll argue that point processes are a good framework when thinking about similar questions

Why are point processes useful in that context?

- ▶ Take the idea that “my model of saliency maps will help me predict where people fixate”
 - ▶ How exactly?
 - ▶ We need a quantitative version of that statement
 - ▶ Plenty of confusion in the literature as to how we make this quantitative, exactly.
- ▶ I'll argue that point processes are a good framework when thinking about similar questions
 - ▶ for example, it's quite clear what we mean when we use a point process for prediction

Why are point processes useful in that context?

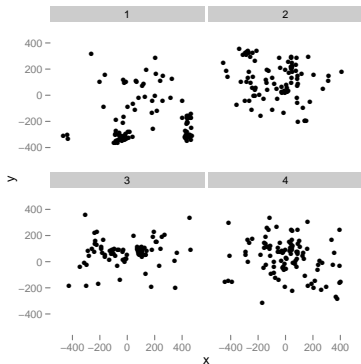
- ▶ Take the idea that “my model of saliency maps will help me predict where people fixate”
 - ▶ How exactly?
 - ▶ We need a quantitative version of that statement
 - ▶ Plenty of confusion in the literature as to how we make this quantitative, exactly.
- ▶ I'll argue that point processes are a good framework when thinking about similar questions
 - ▶ for example, it's quite clear what we mean when we use a point process for prediction
- ▶ Importantly, point processes let us bring many ideas and insights from regression analysis over to spatial data

Point processes

- ▶ Given a spatial domain Ω (e.g. the monitor), a point process is a distribution over finite subsets of Ω .

Point processes

- ▶ Given a spatial domain Ω (e.g. the monitor), a point process is a distribution over finite subsets of Ω .
- ▶ In other words, it generates random point clouds over Ω .
Fixations are exactly that.



Inhomogeneous Poisson processes

- ▶ Poisson processes are the simplest kind of point processes: let S be a realisation from a PP. For all (measurable) $\mathcal{A} \subseteq \Omega$, the number of points of S in \mathcal{A} follows a Poisson distribution with intensity $\lambda_{\mathcal{A}}$.

Inhomogeneous Poisson processes

- ▶ Poisson processes are the simplest kind of point processes: let S be a realisation from a PP. For all (measurable) $\mathcal{A} \subseteq \Omega$, the number of points of S in \mathcal{A} follows a Poisson distribution with intensity $\lambda_{\mathcal{A}}$.
- ▶ Such a process can be completely described via its *intensity function* $\lambda(x, y)$:

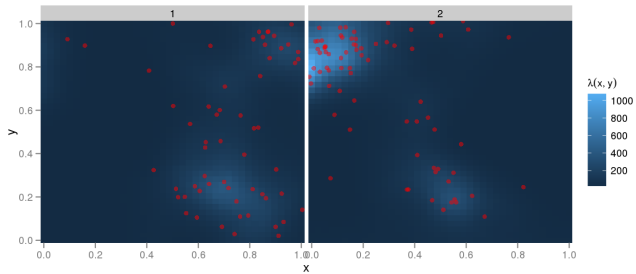
Inhomogeneous Poisson processes

- ▶ Poisson processes are the simplest kind of point processes: let S be a realisation from a PP. For all (measurable) $\mathcal{A} \subseteq \Omega$, the number of points of S in \mathcal{A} follows a Poisson distribution with intensity $\lambda_{\mathcal{A}}$.
- ▶ Such a process can be completely described via its *intensity function* $\lambda(x, y)$:

$$\lambda_{\mathcal{A}} = E(|S \cap \mathcal{A}|) = \int_{\mathcal{A}} \lambda(x, y) dx dy$$

- ▶ The intensity function is the part we are going to model

Two intensity functions



Spatial covariates

- ▶ In the applications we are interested in, the question is how the spatial intensity function relates to *spatial covariates*

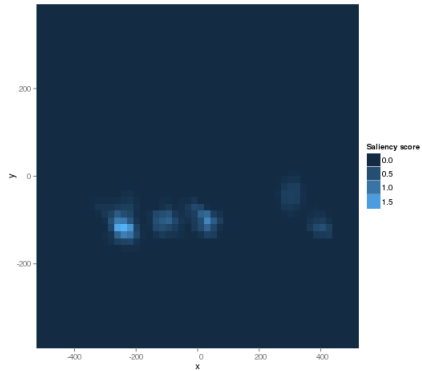
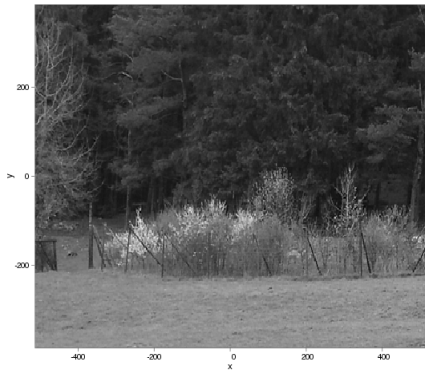
Spatial covariates

- ▶ In the applications we are interested in, the question is how the spatial intensity function relates to *spatial covariates*
- ▶ The obvious spatial covariate here is the saliency map given by the Itti-Koch model.
- ▶ We use the MATLAB Saliency Toolbox (Walther and Koch, 2006).

Spatial covariates

- ▶ In the applications we are interested in, the question is how the spatial intensity function relates to *spatial covariates*
- ▶ The obvious spatial covariate here is the saliency map given by the Itti-Koch model.
- ▶ We use the MATLAB Saliency Toolbox (Walther and Koch, 2006).

Saliency maps



A naive model

$$\log \lambda_i(x, y) = \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ $m_i(x, y)$ is the I&K saliency map.

A naive model

$$\log \lambda_i(x, y) = \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ $m_i(x, y)$ is the I&K saliency map.
- ▶ β_i coefficient governs the relationship between I&K saliency and intensity.

A naive model

$$\log \lambda_i(x, y) = \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ $m_i(x, y)$ is the I&K saliency map.
- ▶ β_i coefficient governs the relationship between I&K saliency and intensity.
- ▶ i indexes the image

A naive model

$$\log \lambda_i(x, y) = \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ $m_i(x, y)$ is the I&K saliency map.
- ▶ β_i coefficient governs the relationship between I&K saliency and intensity.
- ▶ i indexes the image

Fitting the model in R

- ▶ One line in R

```
res <- mppm(ppdat[1:100], ~ image:salmap + image,  
covar.ind=list(salmap=salmap))
```

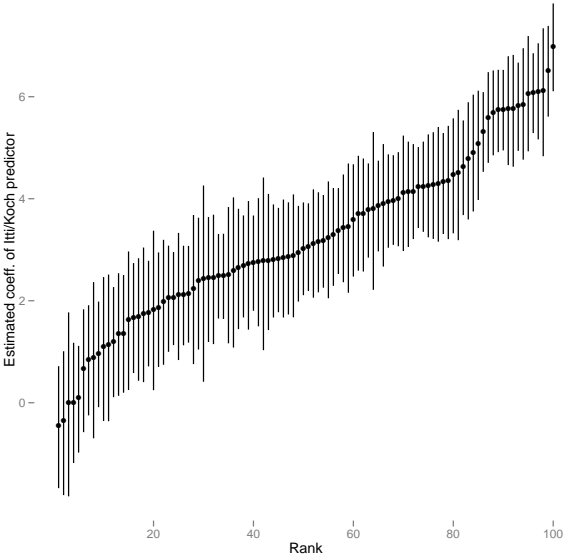
Fitting the model in R

- ▶ One line in R

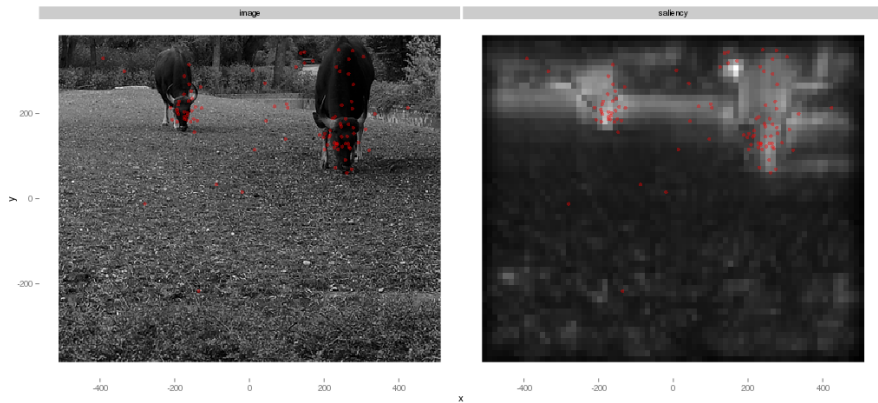
```
res <- mppm(ppdat[1:100], ~ image:salmap + image,  
covar.ind=list(salmap=salmap))
```

- ▶ Estimates the β 's from data using approximate Bayesian inference (INLA, Rue et al., 2009), for the first 100 images in the set.

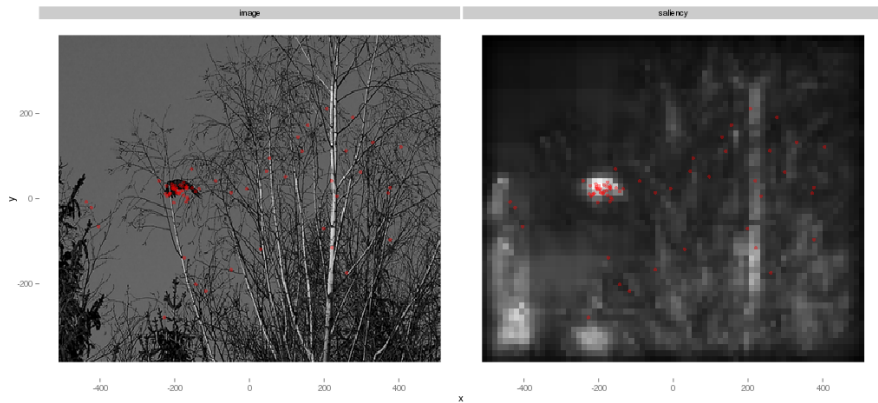
Estimated coefficients



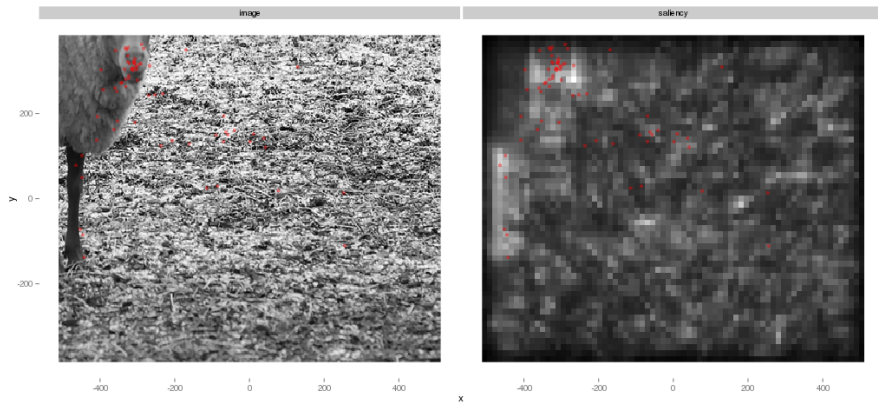
The best



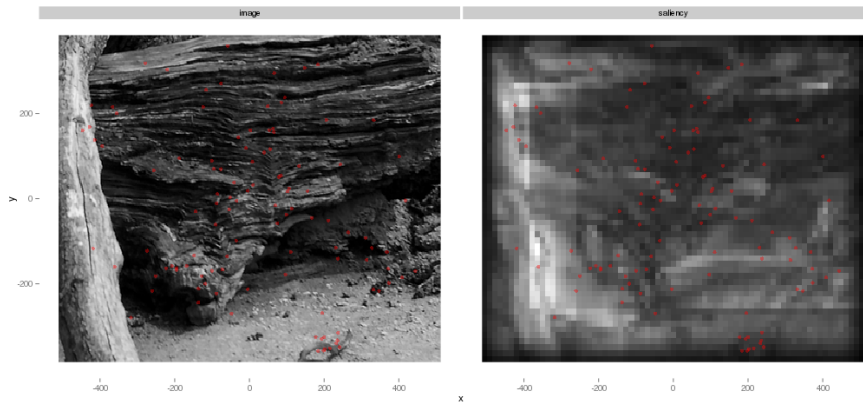
The best



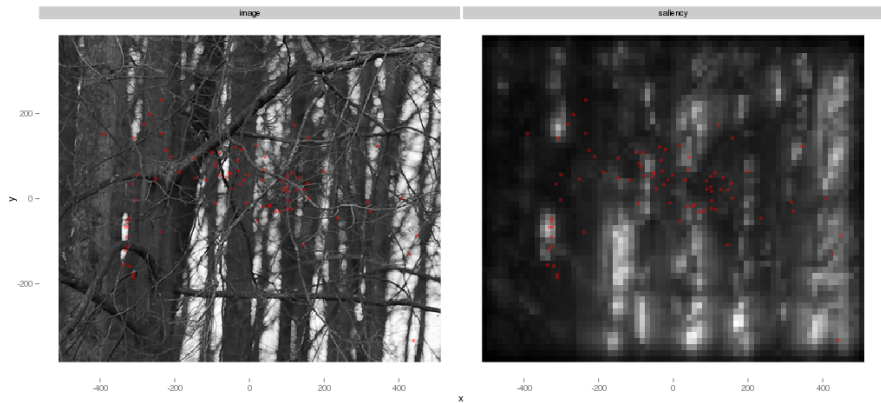
The best



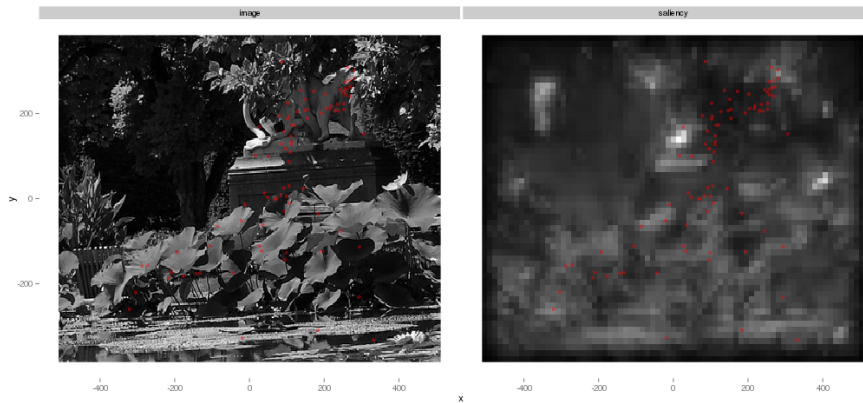
The worst



The worst

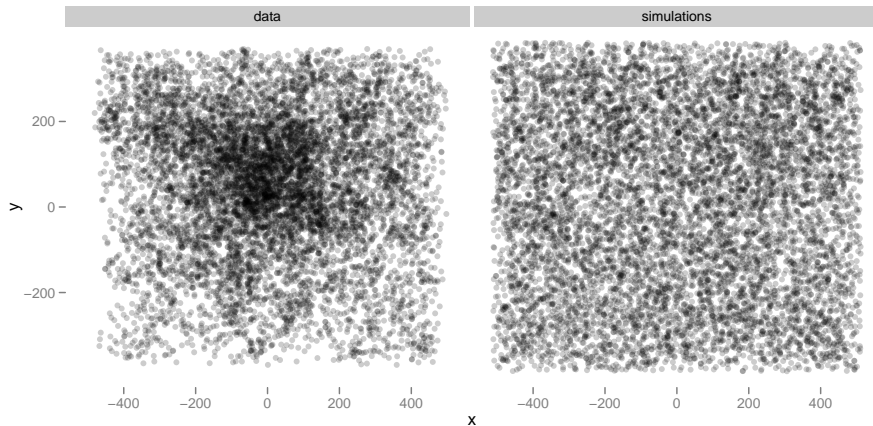


The worst



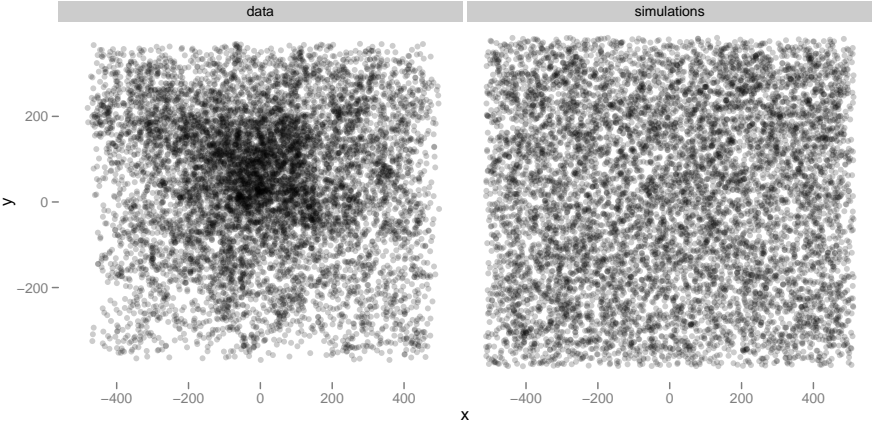
Less obvious failures

- ▶ Some problems become visible when checking marginal statistics against model fit.



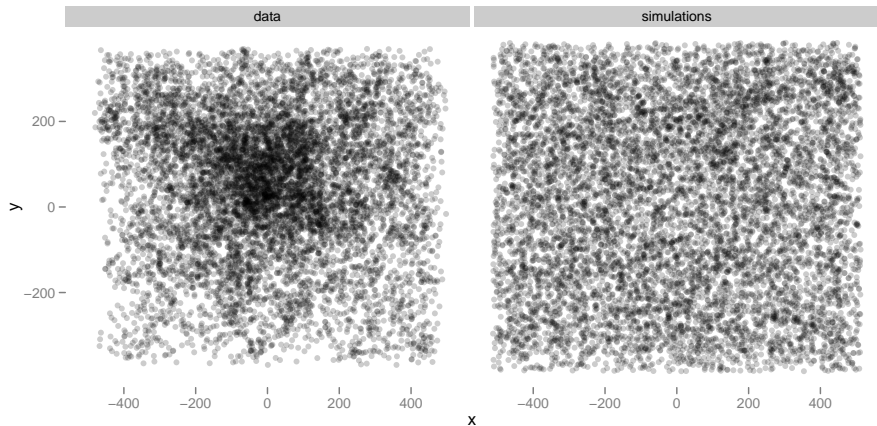
Less obvious failures

- ▶ Some problems become visible when checking marginal statistics against model fit.



Less obvious failures

- ▶ Some problems become visible when checking marginal statistics against model fit.



A better model

- ▶ Among other problems, we have an instance of the “centrality bias” (Tatler, 2007): people like to fixate around the center regardless of what’s there.

A better model

- ▶ Among other problems, we have an instance of the “centrality bias” (Tatler, 2007): people like to fixate around the center regardless of what’s there.
- ▶ Fortunately, there’s a fairly easy fix:

$$\log \lambda_i(x, y) = g(x, y) + \beta_i \times m_i(x, y) + \alpha_i$$

A better model

- ▶ Among other problems, we have an instance of the “centrality bias” (Tatler, 2007): people like to fixate around the center regardless of what’s there.
- ▶ Fortunately, there’s a fairly easy fix:

$$\log \lambda_i(x, y) = g(x, y) + \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ $g(x, y)$ will capture spatial biases.

A better model

- ▶ Among other problems, we have an instance of the “centrality bias” (Tatler, 2007): people like to fixate around the center regardless of what’s there.
- ▶ Fortunately, there’s a fairly easy fix:

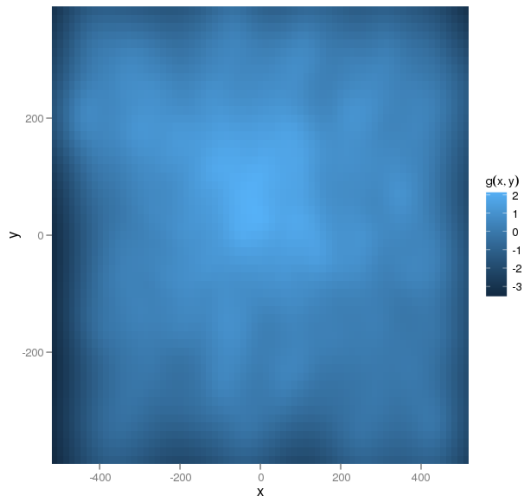
$$\log \lambda_i(x, y) = g(x, y) + \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ $g(x, y)$ will capture spatial biases.
- ▶ We put a Gauss-Markov process prior on $g(x, y)$ to enforce spatial smoothness (Rue and Held, 2005).

The new code

```
res <- mppm(ppdat, ~ s(x,y) + image:salmap + image,  
covar.ind=list(salmap=salmap))
```

Estimated spatial bias



The prediction problem

- ▶ For engineering applications you need to be able to do predictions on new images.
- ▶ Problem: the covariate given by the Itti&Koch model is very helpful on some images and not at all in others

The prediction problem

- ▶ For engineering applications you need to be able to do predictions on new images.
- ▶ Problem: the covariate given by the Itti&Koch model is very helpful on some images and not at all in others

$$\log \lambda_i(x, y) = g(x, y) + \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ For the first 100 images $\hat{\beta}_i$ varies between -2 and 6.

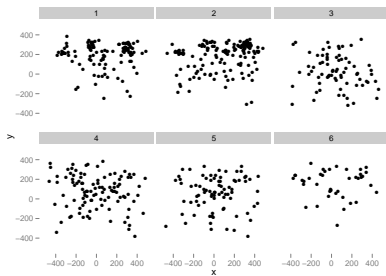
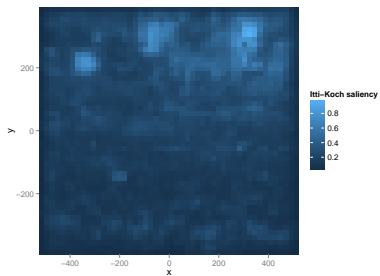
The prediction problem

- ▶ For engineering applications you need to be able to do predictions on new images.
- ▶ Problem: the covariate given by the Itti&Koch model is very helpful on some images and not at all in others

$$\log \lambda_i(x, y) = g(x, y) + \beta_i \times m_i(x, y) + \alpha_i$$

- ▶ For the first 100 images $\hat{\beta}_i$ varies between -2 and 6.
- ▶ When this uncertainty is taken into account predictions become really, really vague...

Some simulations



Conclusions so far

- ▶ We started out with two questions.

Conclusions so far

▶ We started out with two questions.

1. How well does the I&K model predict the data?

Conclusions so far

- ▶ We started out with two questions.
1. How well does the I&K model predict the data?
 - ▶ Answer: has some predictive value on average, but variance across images makes it difficult to use

Conclusions so far

- ▶ We started out with two questions.
 1. How well does the I&K model predict the data?
 - ▶ Answer: has some predictive value on average, but variance across images makes it difficult to use
 2. Can we improve on it?

Conclusions so far

- ▶ We started out with two questions.
 1. How well does the I&K model predict the data?
 - ▶ Answer: has some predictive value on average, but variance across images makes it difficult to use
 2. Can we improve on it?
 - ▶ Yes, one possibility is to include terms capturing spatial biases. There are naturally many other things one could do.

Other data

- ▶ Models I have talked about today are part of a larger class I call “multiple point processes”

Other data

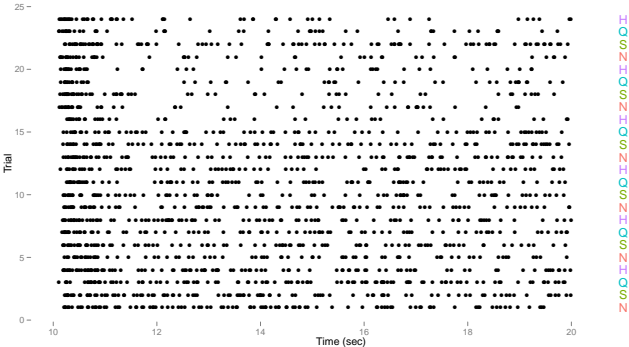
- ▶ Models I have talked about today are part of a larger class I call “multiple point processes”
- ▶ Accordingly the package is called *mpp*

Other data

- ▶ Models I have talked about today are part of a larger class I call “multiple point processes”
- ▶ Accordingly the package is called *mpp*
- ▶ Not just for eye movements

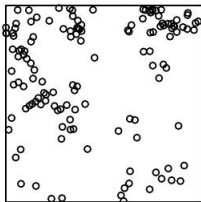
Neural data

- ▶ Models with very similar structure arise in neural data analysis.
- ▶ DiLorenzo & Victor (2003):

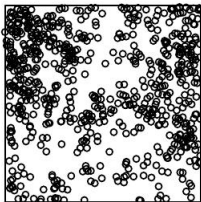


Traditional spatial statistics data

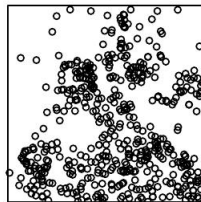
blackoak



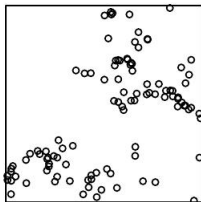
hickory



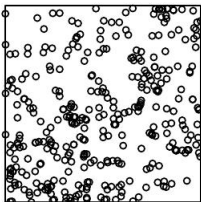
maple



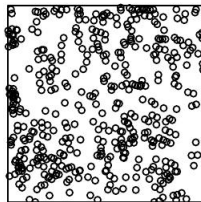
misc



redoak



whiteoak



Conclusions, future work, etc.

- ▶ It also does *temporal point processes*
 - ▶ Spike trains!

Conclusions, future work, etc.

- ▶ Right now *mpp* uses INLA as a back-end, but ideally I'd like to have:
 - ▶ Penalised likelihood using *mgcv*
 - ▶ MCMC using ??? (nothing I've tried works well enough)

Conclusions, future work, etc.

- ▶ Right now *mpp* uses INLA as a back-end, but ideally I'd like to have:
 - ▶ Penalised likelihood using *mgcv*
 - ▶ MCMC using ??? (nothing I've tried works well enough)
- ▶ There's a very nice extension to non-parametric Markov chains

Conclusions, future work, etc.

- ▶ Right now *mpp* uses INLA as a back-end, but ideally I'd like to have:
 - ▶ Penalised likelihood using *mgcv*
 - ▶ MCMC using ??? (nothing I've tried works well enough)
- ▶ There's a very nice extension to non-parametric Markov chains
- ▶ There's a tech report coming up: should be on arXiv next week

Conclusions, future work, etc.

- ▶ Right now *mpp* uses INLA as a back-end, but ideally I'd like to have:
 - ▶ Penalised likelihood using *mgcv*
 - ▶ MCMC using ??? (nothing I've tried works well enough)
- ▶ There's a very nice extension to non-parametric Markov chains
- ▶ There's a tech report coming up: should be on arXiv next week
- ▶ Inference in large and/or high-resolution models is still very much an unsolved problem.

Collaborators

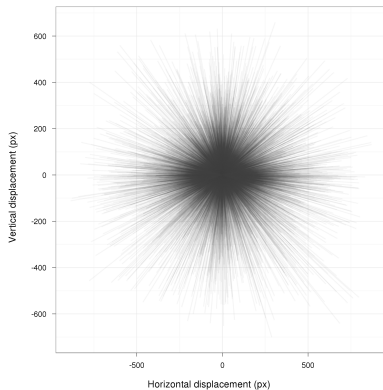
- ▶ Hans Trukenbrod & Ralf Engbert (U Potsdam)
- ▶ Felix Wichmann (U Tübingen)

References

- Berman, M. and Turner, T. R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics*, 41(1):31–38.
- Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns (Statistics in Practice)*. Wiley-Interscience, 1 edition.
- Kienzle, W., Franz, M. O., Schölkopf, B., and Wichmann, F. A. (2009). Center-surround patterns emerge as optimal predictors for human saccade targets. *Journal of vision*, 9(5).
- Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)*. Chapman and Hall/CRC, 1 edition.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):1–17.
- Walther, D. and Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395–1407.

Extensions

- ▶ Eye movements shown strong temporal dependencies: on average the eye do not move very much.



Spatiotemporal models

- ▶ One can use a model such as:

$$\log p_i(\mathbf{s}_t | \mathbf{s}_{t-1}) = g(\mathbf{s}_t) + d(|\mathbf{s}_t - \mathbf{s}_{t-1}|) + f_i(\mathbf{s}_t) + c$$

Spatiotemporal models

- ▶ One can use a model such as:

$$\log p_i(\mathbf{s}_t | \mathbf{s}_{t-1}) = g(\mathbf{s}_t) + d(|\mathbf{s}_t - \mathbf{s}_{t-1}|) + f_i(\mathbf{s}_t) + c$$

- ▶ This is a non-parametric Markov model

Spatiotemporal models

- ▶ One can use a model such as:

$$\log p_i(\mathbf{s}_t | \mathbf{s}_{t-1}) = g(\mathbf{s}_t) + d(|\mathbf{s}_t - \mathbf{s}_{t-1}|) + f_i(\mathbf{s}_t) + c$$

- ▶ This is a non-parametric Markov model
- ▶ Interesting because it can be used to infer *intrinsic saliency* in an image

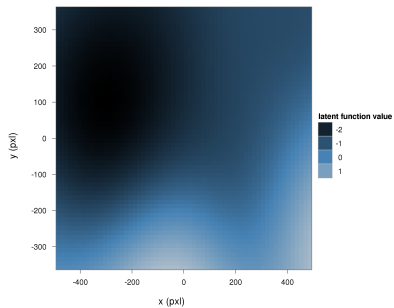
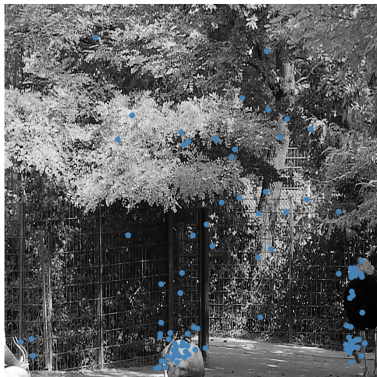
Inference on the non-parametric Markov model

- ▶ An extension of the classical multinomial-Poisson transform to the continuous case shows that, for all practical purposes, we can treat our n fixations as n independent Poisson point processes with only one point each.

Inference on the non-parametric Markov model

- ▶ An extension of the classical multinomial-Poisson transform to the continuous case shows that, for all practical purposes, we can treat our n fixations as n independent Poisson point processes with only one point each.

Results



The IPP likelihood and its approximation

$$\log p(S|\lambda(\cdot)) = \sum_{i=1}^{|S|} \log \lambda(s_i) - \int_{\Omega} \lambda(x) dx$$

- ▶ The problem is with the integral over the domain: rarely closed form.
- ▶ Many similar solutions appeared independently, I think Berman and Turner (1992) were the first
- ▶ Use a grid of values x_1, \dots, x_k :
 - ▶ $\int_{\Omega} \lambda(x) dx$ can be approximated by quadrature
 - ▶ $\log \lambda(s_i)$ can be approximated by interpolation (many variants)