

Some Examples of Random Process Environmental Data Analysis

David R. Brillinger

1. Introduction

Data of process type are now routinely collected and analyzed in the environmental sciences. This is a consequence, in part, of today's general availability of sophisticated computing, storage, display and analysis equipment. At the same time stochastic models have been developed that take detailed note of the special characteristics of such data and hence allow more appropriate and efficient analyses to be carried through. The problems can be difficult, but often an approach is suggested by basic scientific background and the parameters have physical interpretations. Recognizing a process type is an important step along the way to its analysis. The goal of this work is to bring out some basic ideas by presenting a number of elementary examples of random process data analysis.

The work proceeds by describing some basic types of stochastic processes and then presenting some techniques for addressing general problems arising. The emphasis is on processes, their characteristics and understanding their nature by descriptive statistics and elementary analyses rather than by developing background theory. By presenting examples, from different fields, and doing so in comparative fashion the intention is to bring out both similarities and differences. The examples have differing goals.

Concern will be with how the data might be presented visually and described analytically. The next section presents a few basic formal concepts. Section 3 is concerned with temporal point and marked point

processes and an application to a risk assessment problem in space science. Section 4 is concerned with a count-valued time series relating to concerns with childbirth risk. Section 5 focuses on spatial-temporal processes with an example from neuroscience. Section 6 focuses on particle processes with an example from marine biology and the beginnings of an example from pest management. Finally there are some general remarks and discussion.

Techniques highlighted include: plotting the data, likelihood analysis, the EM method, generalized linear modeling, Fourier inference and state space modelling.

2. Some Basic Concepts and Methods

A classical and effective approach for addressing a broad variety of environmental problems is to view the data that have come to hand as part of a realization of a stochastic process. In simplest terms a *random process* is a family of random variables indexed by a label. In the present work the label will refer to time or space-time. The data of concern may be real-valued, vector-valued, categorical-valued, or generalized function-valued amongst other possibilities.

A random process may be described as a family of jointly distributed random variables. The values it takes on and the character of the index labelling the members of the family are what provide special features. For example a *temporal point process*, referring to the occurrence times of some event of interest, might be described by providing the joint distributions of the count-valued random variables $N(A_1), \dots, N(A_k)$ where A_1, \dots, A_k , k in $Z = \{0, \pm 1, \pm 2, \dots\}$, refer to any Borel subsets of $R = (-\infty, \infty)$ and $N(A)$ is the number of occurrences of the event in the set

A. Of course the distributions must be consistent for the process to be well-defined. In the *stationary* case these distributions will be invariant under translations in time. Stationarity is basic to the definition of important parameters describing processes and to the derivation of the statistical properties of quantities computed from process data. In Section 6 an example of points distributed in both the plane and time is considered. Then the Borel subsets are contained in R^3 .

Two general approaches will be made use of in the analyses presented. In the *Method of Moments* basic use is made of moments and cumulants to define parameters of importance and to develop properties of polynomial-type statistics based on data at hand. In *Likelihood Analysis* a serious attempt is made to set down a full model and thereby obtain efficient procedures. Bayesians would recommend multiplying the likelihood further by a prior distribution.

3. Point Processes and Marked Point Processes

3.1. Background

A *temporal point process* is a collection of occurrence times of events, $\{\tau_j\}$, supposed distinct and ordered by $\tau_j < \tau_{j+1}$, j in Z and τ_j in R .

Practically, it is often useful to describe a point process via its *conditional intensity* function. Among those introducing this approach into statistics were Cox and Lewis (1972), Rubin (1972) and Snyder (1975). To describe the conditional intensity of a temporal point process write

$$N(t) = \# \{ \tau_j \text{ in } [0,t) \} = N[0,t)$$

and $H_t = \{\tau_j \text{ with } \tau_j \leq t\}$. This last is referred to as the *history* of the process up to time t . When it exists, the *conditional intensity*, $\mu(t | H_t)$, is given by

$$Prob \{dN(t) = 1 | H_t\} = \mu(t | H_t)dt$$

with the interpretation that $\mu(t | H_t)$ is the rate of occurrence of events at time t given the history until and including then. With occurrence times $0 \leq \tau_j < T$ and supposing the process distribution to depend on a parameter θ the *likelihood function* is given by

$$\prod_j \mu(\tau_j | H_t, \theta) \exp\left\{-\int_0^T \mu(t | H_t, \theta)dt\right\}$$

It may be used to make inferences concerning θ .

Important characteristics of a stationary point process may sometimes be inferred from an estimate of its *autointensity* function

$$m(u) = Prob \{dN(t+u) = 1 | dN(t) = 0\}dt \quad (3.1)$$

giving the rate at which points occur at lag u after an existing point. It may be estimated directly, see Brillinger (1978).

A *marked temporal point process* is a sequence of pairs $\{(\tau_j, M_j)\}$, with τ_j in R referring to the j -th time and M_j an associated quantity (or mark) at that time. The likelihood function may be based on the probability element

$$Prob \{dN(t) = 1 \text{ and } m < M_{N(t)} < m+dm | H_t\} = v(t, dm | H_t)dt \quad (3.2)$$

see Fishman and Snyder (1976). For example one then has

$$Prob \{no \text{ point in } (t, t+u) \text{ with mark } \geq m | H_t\} = \exp\left\{-\int_t^{t+u} \int_m^\infty v(s, dm | H_s)ds\right\}$$

(3.3)

Writing

$$o(s, dm | H_t) = E \{v(s, dm | H_t)\} \quad (3.4)$$

the probability (3.3) is bounded by

$$\int_t^{t+u} \int_m^\infty o(s, dm | H_t) ds \quad (3.5)$$

see Brillinger (1982).

A common question is whether the temporal and mark variations are statistically independent.

3.2. An Example From Space Science

Astronauts living and working in space are subject to a wide variety of risks of which an important one is that they, or their space craft, may be hit by orbiting debris. To assess this risk NASA sampled the population of orbiting objects, see Committee on Orbital Debris (1995). A narrow radar beam was used to detect and estimate characteristics of debris, data being collected over a number of observation periods when the Haystack telescope was available.

Figure 1a displays, $N(t)$, the cumulative count of times at which pieces of orbital debris, at altitudes between 700 and 1100 km, passed through the field of view of a radar beam for one observation period. For the data graphed 33 pieces were detected passing through in 160.1 minutes. In the stationary case the step function should fluctuate around a straight line, as appears reasonably the case here.

Were the process homogeneous Poisson the intervals would be independent exponentials with the same mean. Figure 1b provides a plot of the points $(Y_{(j)}, j/(n+1))$ where the $Y_{(j)}$ are the order statistics of the intervals, $\tau_{j+1} - \tau_j$, between successive times. In preparing the figure the

data for all the observation periods were employed. The plot would be approximately linear were the intervals homogeneous exponentials. A straight line has been applied to the plot as a reference and the exponential appears a reasonable working hypothesis, but there is a hint of departure.

A point process may be a *renewal* process, that is the intervals independent and identically distributed. Turning to this possibility the presence of serial correlation amongst the intervals is assessed. Given a stretch of values, $Y_j, j=0, \dots, J-1$, the *periodogram* is defined by

$$\frac{1}{2\pi J} \left| \sum Y_j \exp\{-i\lambda j\} \right|^2$$

In the case of independence this statistic will fluctuate about a constant level. Figure 2a provides the average of the periodograms of the sequences of intervals between objects averaging over the observation stretches. Also included on the plot is an estimate of that constant level and approximate 90% confidence interval lines assuming the basic process stationary and mixing. There is little evidence against the assumption of a renewal process for this data set.

The autointensity function (3.1) may also be used to examine the Poisson assumption. In the Poisson case it would be constant at the mean rate of the process. Figure 2b provides an estimate of the square root. The estimate employed merges the data from all observation periods. Approximate 90% confidence limits are indicated by the dotted lines. One sees a suggestion that the intensity is raised at lags .04 to .08, but the suggestion is not strong. A possibility is that the process could be renewal with a non-exponential interval distribution. The square root has been graphed here because in the case of reasonably lengthy stretches of data the sampling fluctuations are approximately constant.

In summary, a homogeneous Poisson process appears a plausible working hypothesis for the point process of passage times of these objects.

3.3 Continuing the Space Science Example

Figure 3a provides the times of passage of the same particles as in Figure 1a, but now the estimated altitudes of the particles are also indicated by the heights of the vertical lines. In collecting the data the sizes of the objects were also estimated, by the so-called radar cross section. This measure has an (imperfect) connection with the physical size, see Levanon (1988). Figure 3b extends Figure 3a by including the sizes of the particles. The altitude is still indicated by the y -axis height but the sizes of the objects are indicated by the radii of circles. The data here may be viewed as part of a realization of a marked point process with mark $M = (\textit{altitude}, \textit{radar cross section})$.

Questions of interest include whether the sequence of marks $\{M_j\}$ is independent of the sequence of times $\{\tau_j\}$, and whether the sequences of altitudes and sizes are themselves independent and identically distributed (i.i.d.). The first question was raised in the context of earthquake sequences by Vere-Jones (1970). As will be seen below it may be addressed by spectrum analysis. Figures 4a and 4b provide average periodograms of the altitude and size values in the manner of Figure 2a. They both have the character of white noise processes. The estimates are almost totally within the approximate 90% confidence limits.

As mentioned above it is of interest to ask whether the sequence of mark values is independent of the temporal point process. One might wonder for example do larger sized objects tend to follow longer gaps? This question may be addressed via cross-spectral analysis in the case that it appears plausible to assume the inherent process stationary. Figures 4c

and $4d$ are estimates of the coherences of the sequence of intervals, $Y_j = \tau_{j+1} - \tau_j$, with the altitude and size sequences respectively. These figures provide no evidence of substantial dependence.

Having an approximate model one can now use expressions (3.3), (3.4) to estimate some risks of interest. Taking as working model the basic point process to be Poisson of rate μ , the altitude sequence as independently i.i.d. with density $f_A(a)$ and the sizes as further independently i.i.d. with density $f_S(s)$, the intensity function o of (3.2) is given by

$$\mu f_A(a) f_S(s)$$

as is o of (3.4). The quantities appearing here may be estimated simply and thereby bounds such as (3.5) estimated.

The strength of the data analyses presented is that a broad class of alternative possibilities have been considered prior to obtaining a very simple working model.

4. Time Series

4.1. Background

A *time series* is a wiggly line, $\{Y(t)\}$, with $Y(t)$ in R and t in Z or in R . If $Y(t)$ is binary, taking on the values 0, 1, and the 1's are rare then the series Y appears like a temporal point process.

Given expressions for the conditional mass or density functions, such as

$$Prob \{y \leq Y(t+1) < y+dy \mid H_t, \theta\} = p(t+1 \mid H_t, \theta) dy$$

in the case of t in Z , one can express the likelihood as

$$\prod_t p(t+1 | H_t, \theta) \quad (4.1)$$

Here H_t is the history $\{Y(u), u \leq t\}$. The likelihood can be used to make inferences concerning θ .

4.2 Example from Public Health

The United States has a worrying high level of cesarean deliveries compared to most developed nations, Clarke and Taffel (1995). This shows itself in a substantially reduced number of babies born on weekends. Besides cesareans, the increased proportion of weekday births may be due to the number of births that are induced. Both of these are causes of concern because of increased maternal and infant health risks, *ibid*. The preceding authors list average numbers of births each day of the week for the whole USA from Sunday to Saturday as 8754, 11398, 12333, 11957, 11895, 11957, 9420 respectively for the year 1992.

In this section the dependence of delivery day on the day of the week is studied for the city of Toronto in 1986. Figure 5a graphs the number of births for each day of the year. One notes a rapid oscillation and a bowing up in the middle. The bowing corresponds to more births in the summer. The apparent dependence of the number of births on the time of year, in addition to day of the week, is something that has been noted various times before. The smooth curve added is an estimate of an underlying slowly changing rate as obtained by the function `lowess()` of the statistical package S, see Becker, Chambers and Wilks (1988), Cleveland et al (1992). Figure 5b provides parallel box plots of the birth counts for each day of the week. The lower counts for the first and last days correspond to the reduced number of Saturday and Sunday births. The weekday-weekend phenomenon mentioned is quite pronounced here.

The series values, $Y(t)$, are actually counts and so it appears sensible to employ a model taking some note of this. Consider modelling the count on day t as Poisson with mean

$$\mu_t = \exp\{T_t + S_{\langle t \rangle}\}$$

where T_t is a slowly varying trend component and $S_{\langle t \rangle}$ refers to day of week effect, $\langle t \rangle$ being the day of the week date t falls on. Further consider modelling the successive daily counts as statistically independent. Because of the assumed independence expression (4.1) simplifies to

$$\prod_t \frac{1}{Y(t)!} (\mu_t)^{Y(t)} \exp\{-\mu_t\}$$

This model may be fit directly via the function `gam()` of S, Hastie (1992).

The estimates \hat{T}_t , $\hat{S}_{\langle t \rangle}$ obtained are similar to those suggested by Figure 5. To examine the Poisson assumption the residuals of the fit may be examined for overdispersion. The estimate of the overdispersion parameter is 1.07502 so any overdispersion appears mild.

Figure 6a provides the periodogram of the original count values. The weekly effect is apparent through the presence of the peaks near $1/7$ and $2/7$. The trend shows itself in the higher values near frequency 0. Figure 6b is the periodogram of the Pearson residuals, having removed the estimated trend and day of the week effects. Included on the plot are approximate 95% confidence limits about the independent noise level. There is no strong suggestion of autocorrelation amongst the residuals. The model of independent Poisson counts therefore appears useful.

Were some autocorrelation suggested in the latter plot one could use the function `gam()` of S to include it by approximating the series by 0-1 series and including lagged values in the predictor, see Brillinger and

Segundo (1979). Becker (1986) fits an epidemic process via a conditional binomial generalized linear model.

A model such as the one obtained may be used, for example, to estimate possible changes in costs resulting from fewer elective cesareans.

5. Spatial-temporal Processes

5.1 Background

Spatial-temporal process data may be written $Y(\mathbf{r},t)$ with $\mathbf{r} = (x,y)$ or (x,y,z) and (\mathbf{r},t) in some subset of R^3 or R^4 . One argument, t , has a privileged character. Such data may often be conveniently displayed by a sequence of images, by a video or by spinning a surface.

The process may be available everywhere in a lattice or correspond to irregularly placed points. The latter case corresponds to a spatial-temporal point process $\{(\mathbf{r}_j, t_j) \equiv (x_j, y_j, t_j)\}$. Let H_t denote the history of this process up to and including time t . The conditional intensity is given by

$$Prob \{dN(x,y,t) = 1 \mid H_t\} = v(x,y,t \mid H_t) dx dy dt$$

An example of spatial-temporal point process data will be presented in Section 6.3 . References include: Fishman and Snyder (1976), Vere-Jones and Thomson (1984), Rathburn (1993).

5.2 Example from Neuroscience

The next example concerns the olfactory system, that is the sense of smell. Data were collected of the response of a rabbit's sniffing an odor. The rabbit was conditioned to respond to a particular odor. An array of sensors was applied to the brain above the olfactory bulb and electroencephalograms recorded. The array had 64 sensors laid out in an 8 by 8 lattice, 3.5 mm by 3.5 mm. Bursts between breaths were measured with

values taken 2 ms apart. There were $T = 38$ temporal values recorded for each burst and $J = 12$ replicates. For a description of the experiments see Freeman and Grajski (1987).

The data may be written $[Y_j(x,y,t)]$ with (x,y) location, t time and j replicate, $x,y = 1,\dots,8$, $t = 1,\dots,38$, $j = 1,\dots,12$. Figure 7 shows the data of the first replicate. One notes oscillations at a possibly common frequency, with the amplitude of the oscillations varying with position in the array.

The analysis presented focuses on the separation of space and time variation. In spatial-temporal circumstances separation of variables is an important analytic technique.

For the j -th replicate, consider the model

$$Y_j(x,y,t) = a(x,y)\gamma_j(t) + \varepsilon_j(x,y,t) \quad (5.1)$$

with $a(\cdot)$, fixed and standardized by $\sum a(x,y)^2 = 1$, with the $\gamma_j(\cdot)$ independent stationary time series of common power spectrum $f(\cdot)$ and with the $\varepsilon_j(\cdot)$ independent white noise processes of variance σ^2 .

The relationship (5.1) is a form of random effects model. In the case that the γ_j , ε_j are Gaussian one can consider estimating the unknowns by maximum likelihood. This fitting is conveniently carried out in the frequency domain employing the EM method (Dempster et al (1977)). The steps are: first Fourier transform with respect to t to obtain

$$Y_j(x,y,\lambda) = a(x,y)\Gamma_j(\lambda) + E_j(x,y,\lambda)$$

with λ taking the values $2\pi j/38$. Then having some initial values compute

$$\hat{a}(x,y) = \sum_{j,\lambda} Y_j(x,y,\lambda)\overline{\Gamma_j(\lambda)}/C$$

followed by

$$\hat{\Gamma}_j(\lambda) = [\sum_{x,y} Y_j(x,y,\lambda)a(x,y)]f(\lambda)/(f(\lambda) + \sigma^2)$$

and

$$\hat{f}(\lambda) = \sum_j |\Gamma_j(\lambda)|^{2/J}$$

$$\hat{\sigma}^2 = \sum_{j,x,y,\lambda} |Y_j(x,y,\lambda) - a(x,y)\Gamma_j(\lambda)|^{2/JXY\Lambda}$$

respectively. Finally iterate to convergence. The divisor C is chosen so that $\sum \hat{a}(x,y)^2 = 1$, while $J = 12$, $\Lambda = 19$ and $X, Y = 8$.

Figures 8 and 9 show the results of the computations. Figure 8 displays the fitted spatial function, $\hat{a}(x,y)$ in both perspective and contour fashion. It shows an apparent focus of activity. Figure 9 contains the twelve estimated time series components, $\hat{\gamma}_j(t)$. The estimated replicate time series show oscillations, as was to be anticipated from Figure 7. The amplitudes do vary noticeably with replicate.

Figure 10 shows the residual series $Y(x,y,t) - \hat{a}(x,y)\hat{\gamma}_1(t)$, for the first replicate on the same scale as Figure 7. One sees the amplitudes to be much reduced and the series to be noisier.

Following a classic approximation one can act as if the empirical Fourier transform values E_j are approximately Gaussian, with values at the Fourier frequencies independent, but the assumption of Gaussian γ is basic to the maximum likelihood analysis presented.

This type of work may be seen as establishing base values preparatory to seeking possible changes from base values resulting from some treatment.

6. Particle Processes

6.1 Background

A *particle process* represents the path or trajectory of an object moving along a line, around in a plane or about in space. In the case of the plane the trajectory may be represented by $(X(t), Y(t))$ where $X(t)$ and $Y(t)$ give the x - y coordinates of the particle's location at time t . The particle could be meandering or pole-seeking. The representation $(X(t), Y(t))$ is that of a bivariate time series, but the conceptualization of the problem is often quite different.

An example on a grand scale is provided in Eddy and Que (1995) where there is discussion of how to display and analyse aircraft flights over the continental United States. The process is a collection of paths, $(X_j(t), Y_j(t), Z_j(t))$, in R^3 .

6.2 An Example from Ecology

Next an example from ecology is presented - the migration path of an elephant seal. These animals were near extinct at the turn of the century so there is a societal need to learn more about their behavior. Figure 11 graphs the path of one animal as an example. The animal starts from the Channel Islands off Santa Barbara, California, proceeds to the northwest and then returns. A great circle route has been added to the figure for reference. This animal seems to know surprisingly well where she is going. The problem of how to describe such paths is of interest.

In Brillinger and Stewart (1998) a pole seeking model on the sphere is considered. Suppose θ , ϕ , δ respectively denote longitude, colatitude and speed in a coordinate system such that the animal is traveling to the North Pole. Since the method of estimating the noontime location is quite indirect, (based on times of sunrise, noon, sunset recovered when the animal returns) there is measurement error involved. The equations set

down in Brillinger and Stewart (1998) have the form

$$\theta_t' = \theta_t + \tau \varepsilon_t' \quad (6.1)$$

$$\phi_t' = \phi_t + \tau \gamma_t' / \sin \theta_t' \quad (6.2)$$

$$\theta_{t+1} - \theta_t = \frac{\sigma^2}{2 \tan \theta_t} - \delta + \sigma \varepsilon_{t+1} \quad (6.3)$$

$$\phi_{t+1} - \phi_t = \frac{\sigma}{\sin \theta_t} \gamma_{t+1} \quad (6.4)$$

with $\varepsilon, \gamma, \varepsilon', \gamma'$ unit variance independent Gaussian noise processes. The latter two processes correspond to measurement error.

First the model (6.3-4) for the case of no measurement error is fit by maximum likelihood. The values obtained are:

$$\hat{\delta} = .0112(.0011) \text{radians}$$

$$\hat{\sigma} = .00805 \text{radians}$$

The full model (6.3-4) is a nonlinear state space model, see eg. Harvey (1989). A likelihood may be found based on it once one has an expression for the conditional density, $p(\theta_{t+1}', \phi_{t+1}' | H_t, \delta, \sigma, \tau)$, as in (4.1). This is not directly available but may be estimated by Monte Carlo by generating realizations of the processes θ, ϕ and then averaging.

In the case with measurement error, and supposing the outbound speed is δ while the inbound is δ_1 , the estimates are:

$$\hat{\delta} = .0126(.0001)$$

$$\hat{\delta}_1 = .0109(.0001)$$

$$\hat{\sigma} = .000489(.000004)$$

$$\hat{\tau} = .0175(.0011)$$

all in radians. Now the measurement errors, ε_t' , γ_t' appear dominant, not the foraging movement given by the ε and γ of (6.3-4).

More complex Monte Carlo sampling schemes are available to use here, see eg. Stoffer and Wall (1991), Kitagawa (1996), but were unnecessary because of the small sample size and number of parameters involved.

6.3 Example from Pest Management

This last example is not developed as much as the preceding ones. It is meant to illustrate a data type and the beginnings of model development. It could also have been presented in Section 5.

In 1975 a medfly epidemic took place in the Los Angeles area of California, see Routhier (1977), Hagen et al. (1981). It covered the period September 24 to December 1. Figure 12a plots the locations at which medflies or their larvae were discovered. The map is the area north of Santa Monica and the shaded area is the Pacific Ocean. The incident began in the lower part of the figure in Culver City, and ended with trapings at the lower and upper reaches of the region. Figure 12b graphs the numbers observed each day, with a peak the 37th day. These data are once again spatial-temporal. Figure 12 separates the spatial and temporal variables.

To bring out the joint spatial-temporal character Figure 13 shows the locations of medfly sightings for successive nine day periods. Figure 14 plots the distances from the location of the initial sighting for the successive sightings. One sees sightings at a distance as the epidemic ends. A lowess line has been added as a reference.

The interpretation of this data set is complicated by many things

including: the locations of the traps, the inefficiencies of the traps (often described as low), the timing of visits to the traps and the eradication treatment. Treatment began early in October with application of malathion to host plants. It was soon realized that this approach was too slow to achieve eradication. In November a sterile fly release program began and eventually about 20 million such flies were being released each week. All told 500 million sterile flies were released and some 280,000 of these were trapped in the eight months of the program, see Routhier (1977).

The basic entomological and geographical processes are interesting. The bugs will be spreading by flying, sometimes assisted by the wind. Also the numbers will be increasing rapidly as eggs are laid and become adults. Adults will be dying. If at some time a bug flies near a trap they may be attracted and caught. The data will become available only when a trap is examined. The bug trajectories might be modelled as particle processes, as the elephant seal migration was above. The locations of the traps, the ranges of the traps and the timing of trap examinations all affect the data obtained. The eradication effort will need to be included in the models. The mathematics of the spatial-temporal birth and death process are pertinent, see eg. Cox and Isham (1980).

One problem is how to use such data to estimate the characteristics of the overall population.

7. Other Types of Processes, Data and Techniques

There are various other data sets and associated processes that could well have been discussed in the spirit of the paper. Perhaps foremost are the spatial processes, $Y(\mathbf{r})$ with $\mathbf{r} = (x, y)$ or (x, y, z) . One can also mention line processes, hybrids eg. sampled values, $Y(\tau_j)$, processes on

graphs, trees, shapes, tessellations and other geometric entities. Stoyan et al. (1987) is one reference.

Another topic that might have been presented is the case where the values of a process are discrete, falling into categories. The categories may be ordered, that is the values ordinal.

Difficulties arising in working with data have not been discussed. There are problems with: biased estimates, long range dependence, outliers, missing values, ...

8. Discussion

The goal of this paper has been to present in comparative, parallel fashion examples where the basic data may be seen as part of a realization of a random process. Statistics texts often contain substantial material on descriptive statistics, focusing on numerical quantities and figures separately from any stochastic modelling. In part this has been the approach of the present paper. The classic problems of uncertainty estimation and goodness of fit are ever present, but the paper has not focused on these.

9. Acknowledgements

This paper was prepared with the partial support of NSF Grants DMS-9625774 and DMS-9704739. The author thanks Mark Matney, Walter Freeman and Brent Stewart for providing the data sets of Sections 3, 5.2, 6.2 respectively.

References

- [1] Becker, N. G. (1986). A generalized linear modelling approach to the analysis of data from a single epidemic. Pp. 464-467 in Pacific Statistical Congress (Eds. I. S. Francis, B. F. J. Manly and F. C. Lan). Elsevier, Amsterdam.
- [2] Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). The New S Language. Wadsworth, Pacific Grove.
- [3] Brillinger, D. R. (1978). Comparative aspects of the study of ordinary time series and point processes. Pp. 33-133 in Developments in Statistics, Vol. 1 (Ed. P.R. Krishnaiah). Academic, New York.
- [4] Brillinger, D. R. (1982). Some bounds for seismic risk. Bull. Seismol. Soc. America 72, 1403-1410.
- [5] Brillinger, D. R. and Segundo, J. P. (1979). Empirical examination of the threshold model of neuron firing. Biol. Cybernetics 35, 213-228.
- [6] Brillinger, D. R. and Stewart, B. S. (1998). Elephant seal movements: modelling migration. Canad. J. Statistics, to appear.
- [7] Clarke, S. C. and Taffel, S. (1995). Changes in Cesarean delivery in the United States, 1988 and 1993. Birth 22, 63-67.
- [8] Cleveland, W. S., Grosse, E. and Shyu, W. M. (1992). Local regression models. Pp. 309-376 in Statistical Models in S (Eds. J. M. Chambers and T. J. Hastie). Wadsworth, Pacific Grove.
- [9] Committee on Space Debris (1995). Orbital Debris: a Technical Assessment. National Academy Press, Washington.
- [10] Cox, D. R. and Isham, V. (1980). Point Processes. Chapman and Hall, London.
- [11] Cox, D. R. and Lewis, P. A. W. (1966). The Statistical Analysis of Series of Events. Methuen, London.

- [12] Cox, D. R. and Lewis, P. A. W. (1972). Multivariate point processes. Sixth Berkeley Symp. Math. Stat. Prob., 3, 401-448.
- [13] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J. Royal. Statist. Soc. B 39, 1-22.
- [14] Eddy, W. F. and Que, S. (1995). Dynamic three-dimensional display of U.S. air traffic. J. Comp. Graph. Stat. 4, 261-280.
- [15] Fishman, P. M. and Snyder, D. L. (1976). The statistical analysis of space-time point processes. IEEE Trans. Inf. Theory IT-22, 257-274.
- [16] Freeman, W. J. and Grajski, K. A. (1987). Relation of olfactory EEG to behavior: factor analysis. Behavioral Neuroscience 101, 766-777.
- [16] Hagen, K. S., Allen, W. W. and Tassan, R. L. (1981). Mediterranean fruit fly: the worst may be yet to come. California Agriculture, University of California.
- [17] Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge U. Press, Cambridge.
- [18] Hastie, T. J. (1992). Generalized additive models. Pp. 249-308 in Statistical Models in S (Eds. J. M. Chambers and T. J. Hastie). Wadsworth, Pacific Grove.
- [19] Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian state space models. J. Comp. Graph. Stat. 5, 1-25.
- [20] Levanon, N. (1988). Radar Principles. Wiley, New York.
- [21] Rathburn, S. L. (1993). Modelling marked spatio-temporal point patterns. Bull. ISI, Vol. 55, Book 2, 379-396.
- [22] Routhier, W. (1977). Mediteranium fruit fly. Memorandum, California Dept. of Food and Agriculture.

- [23] Snyder, D. L. (1975). Random Point Processes. Wiley, New York.
- [24] Snyder, D. L. and Miller, M. I. (1991). Random Point Processes in Time and Space. Springer, New York.
- [25] Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter. *J. Amer. Statist. Assoc.* 86, 1024-1033.
- [26] Stoyan, D., Kendall, W. S. and Mecke, J. (1987). Stochastic Geometry and Its Applications. Wiley, New York.
- [27] Vere-Jones, D. (1970). Stochastic models for earthquake occurrence. *J. Royal Stat. Soc. B* 32, 1-62.
- [28] Vere-Jones, D. and Thomson, P. J. (1984). Some aspects of space-time modelling. *Proc. XIIth Int. Biom. Conf. Tokyo*, 265-275.

Figure Legends

Figure 1. The top graph, 1a, plots the cumulative count of debris pieces passing through the field of view of the radar after observing has started for some period. Figure 1b is an exponential probability plot for all the observed intervals between successive objects passing.

Figure 2a is the average of the periodograms for the data of the various observation periods. Also included are approximate 95% marginal confidence limits. Figure 2b is the estimated autointensity function with approximate 95% confidence limits.

Figure 3a represents the times of objects passing through and the corresponding altitude. Figure 3b is as Figure 3a, but now circles are included to represent the sizes of the objects.

Figures 4a and 4b are the averages of the periodograms of the altitudes and sizes, averaging over the available observation periods. Figures 4c and 4d are coherence estimates for the intervals between successive passages with the altitude and size series respectively. The upper null 95% marginal confidence line has been added.

Figure 5a provides the number of births in Toronto 1986 for each day of the year. A smooth lowess curve has been superposed. Figure 5b presents parallel stem-and-leafs split by day of the week.

Figure 6a is the periodogram of the series of Figure 5a. Figure 6b is the periodogram of the residuals having removed an estimated trend and the daily effects.

Figure 7. The electroencephalograms of the first replicate of the experiment for the 8 by 8 array.

Figure 8. The estimated spatial function, $a(x,y)$, of the model (5.1).

Figure 9. The estimated latent series for the 12 replicates.

Figure 10. The residual series having fit the model (5.1).

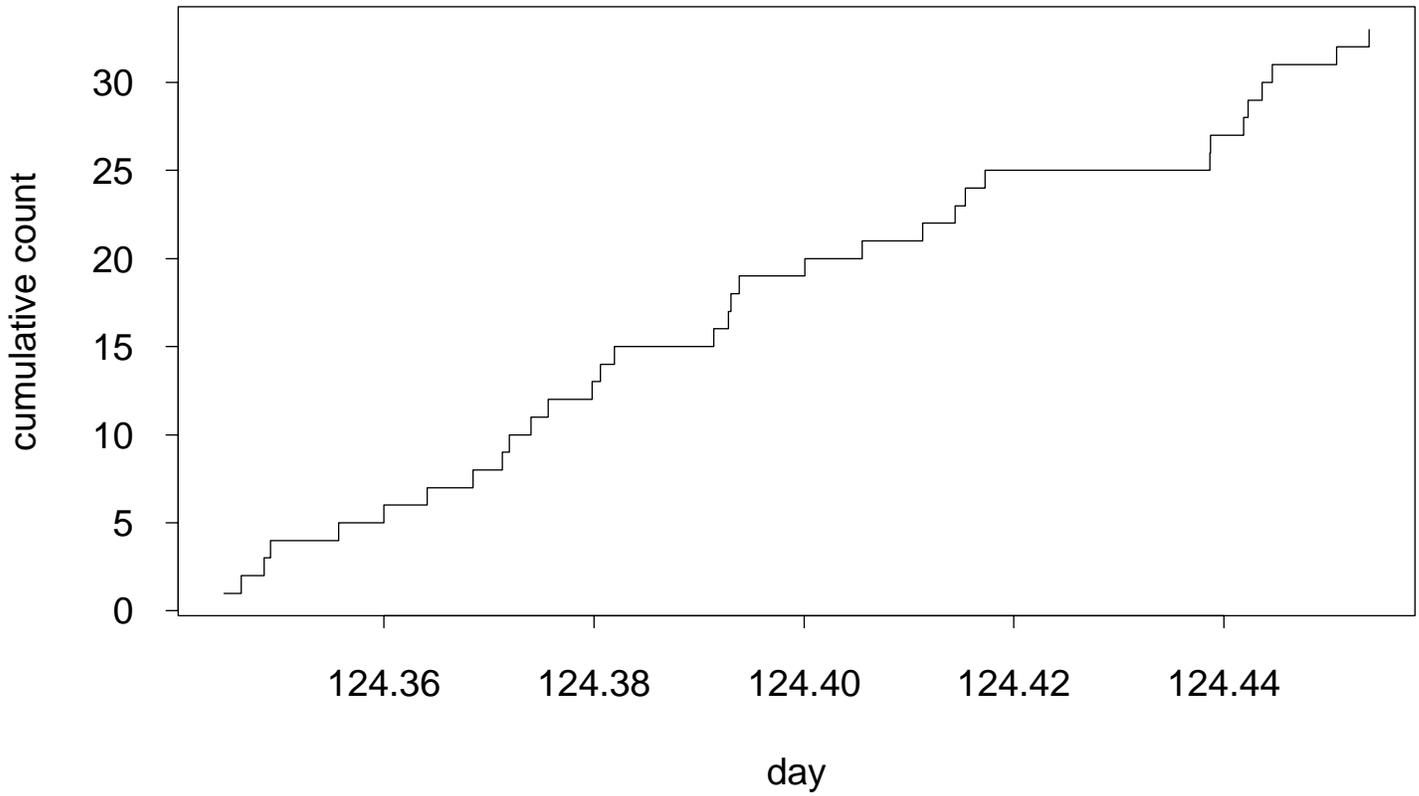
Figure 11. The outbound and inbound tracks of an elephant seal heading into the Northwest Pacific from near Santa Barbara, California.

Figure 12a is a plot of the locations of trappings of medflies, adults and larvae, during an outbreak in 1975. The shaded region is the Pacific Ocean. Figure 12b graphs the counts noted each day during the epidemic.

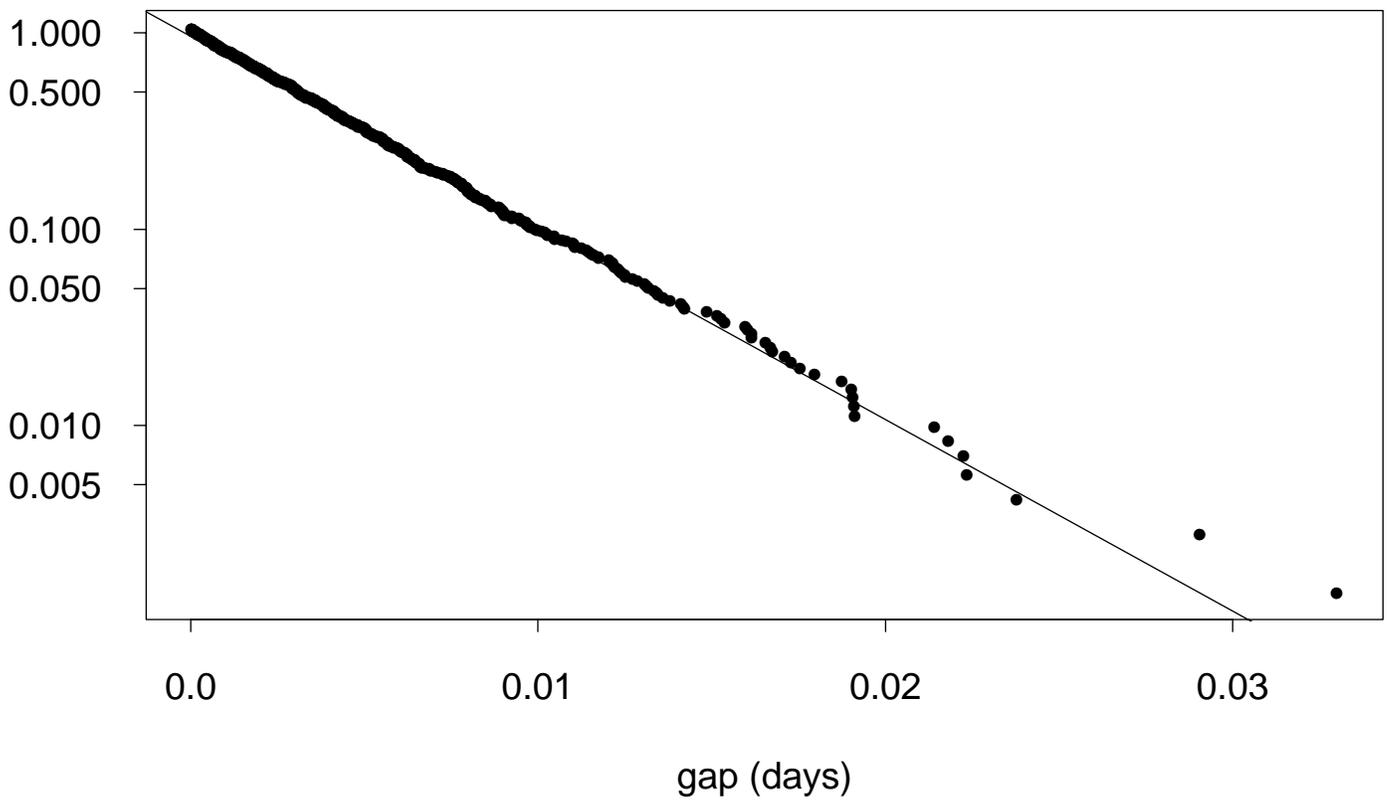
Figure 13. The plot of Figure 12a, but for successive 9 day periods.

Figure 14. The distances of each sighting from the original, for each day.

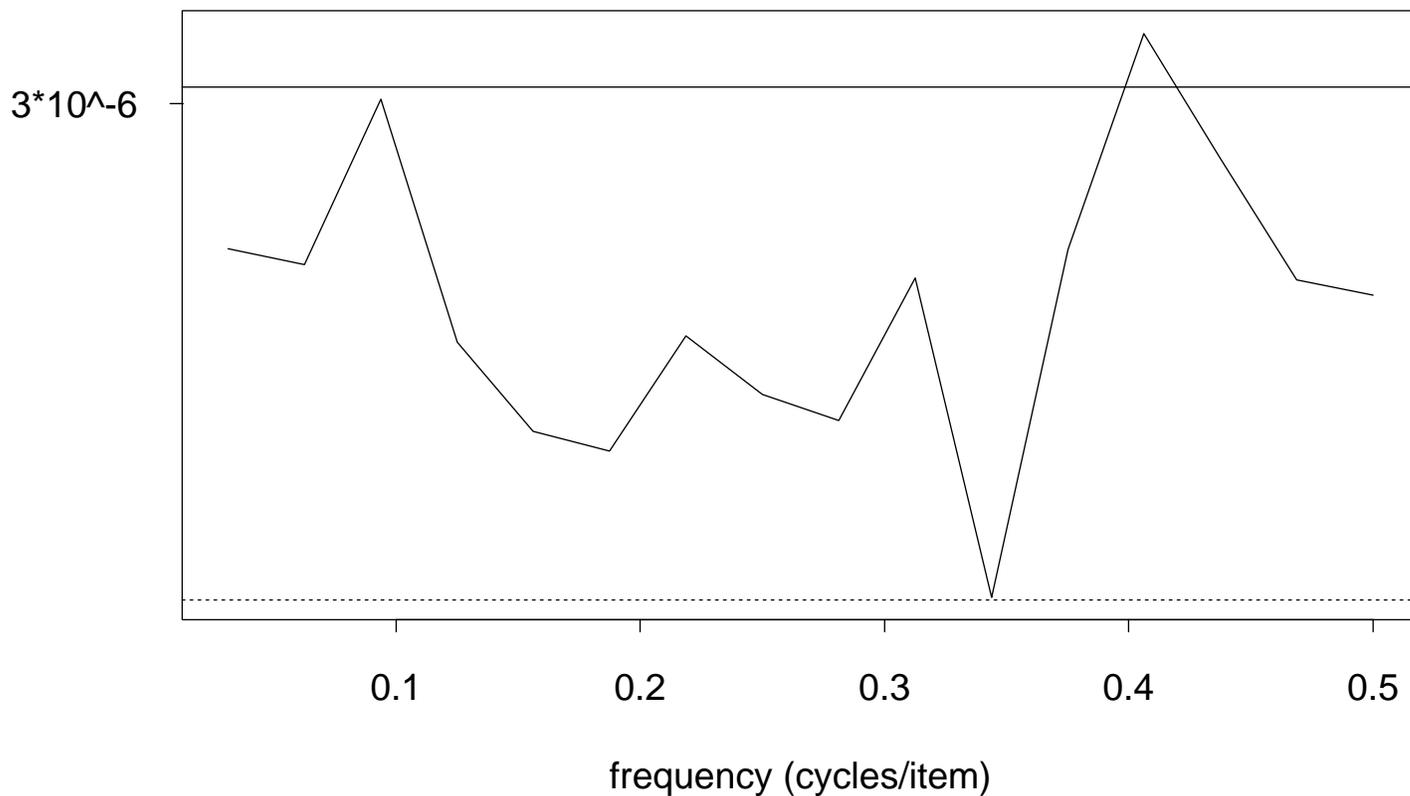
Cumulative count of debris pieces



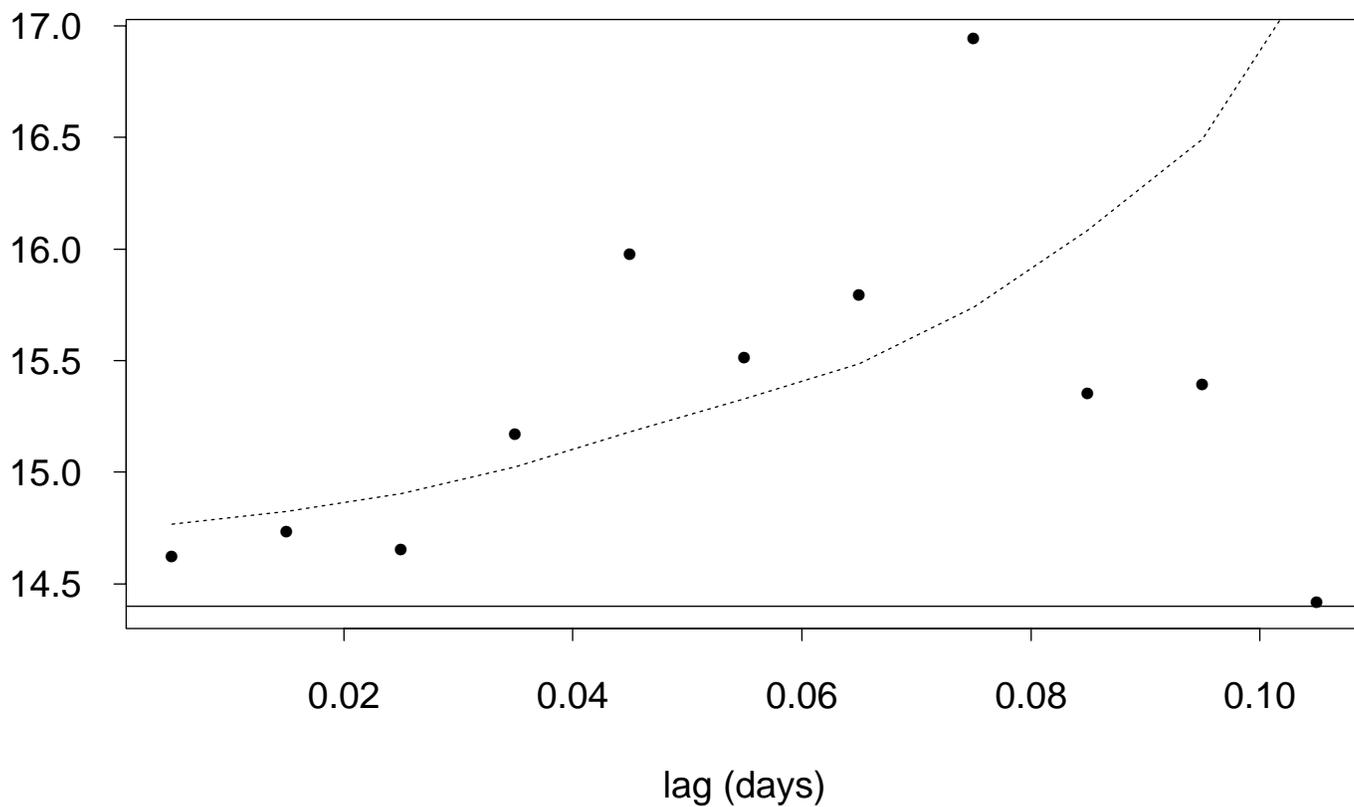
Proportion greater than given gap



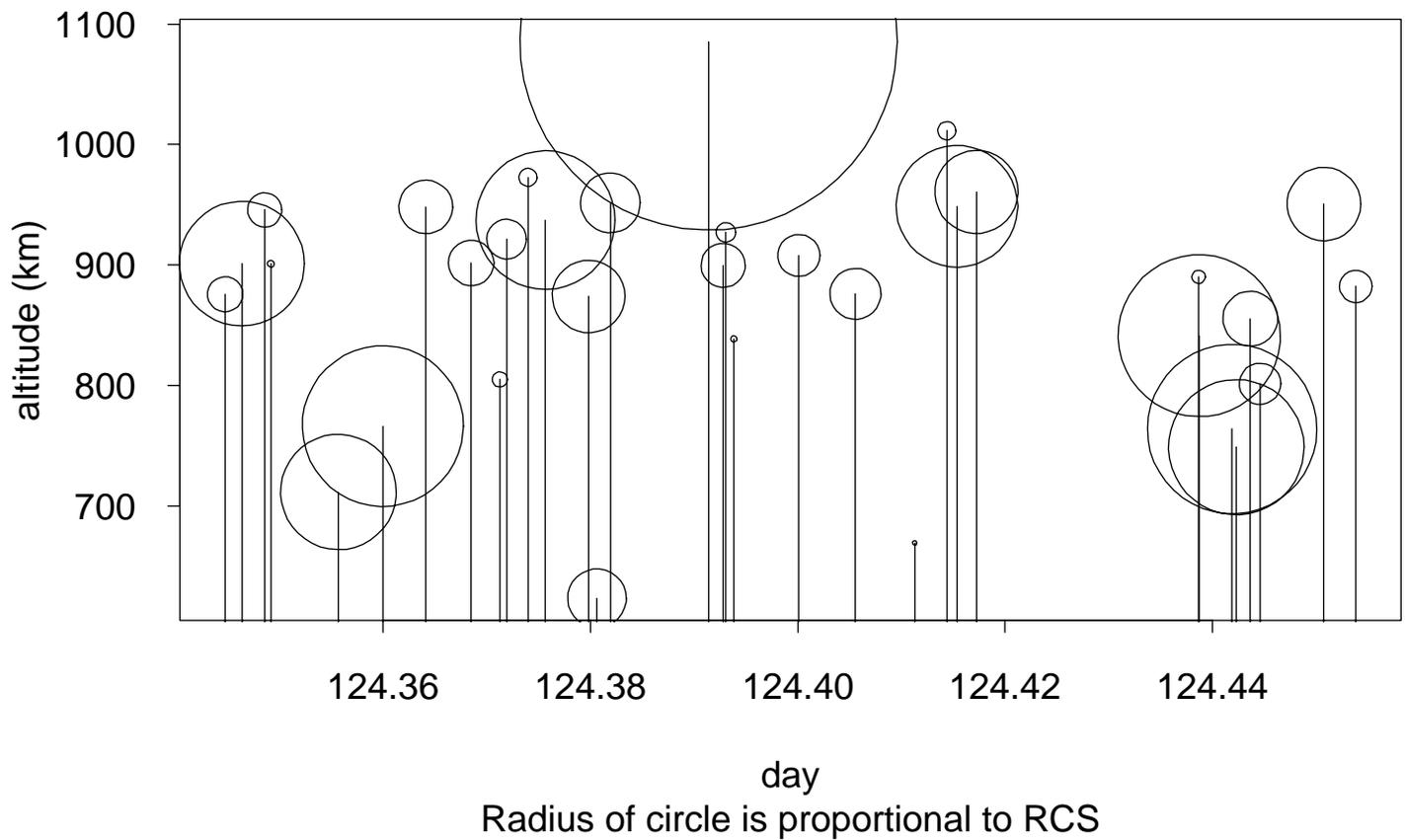
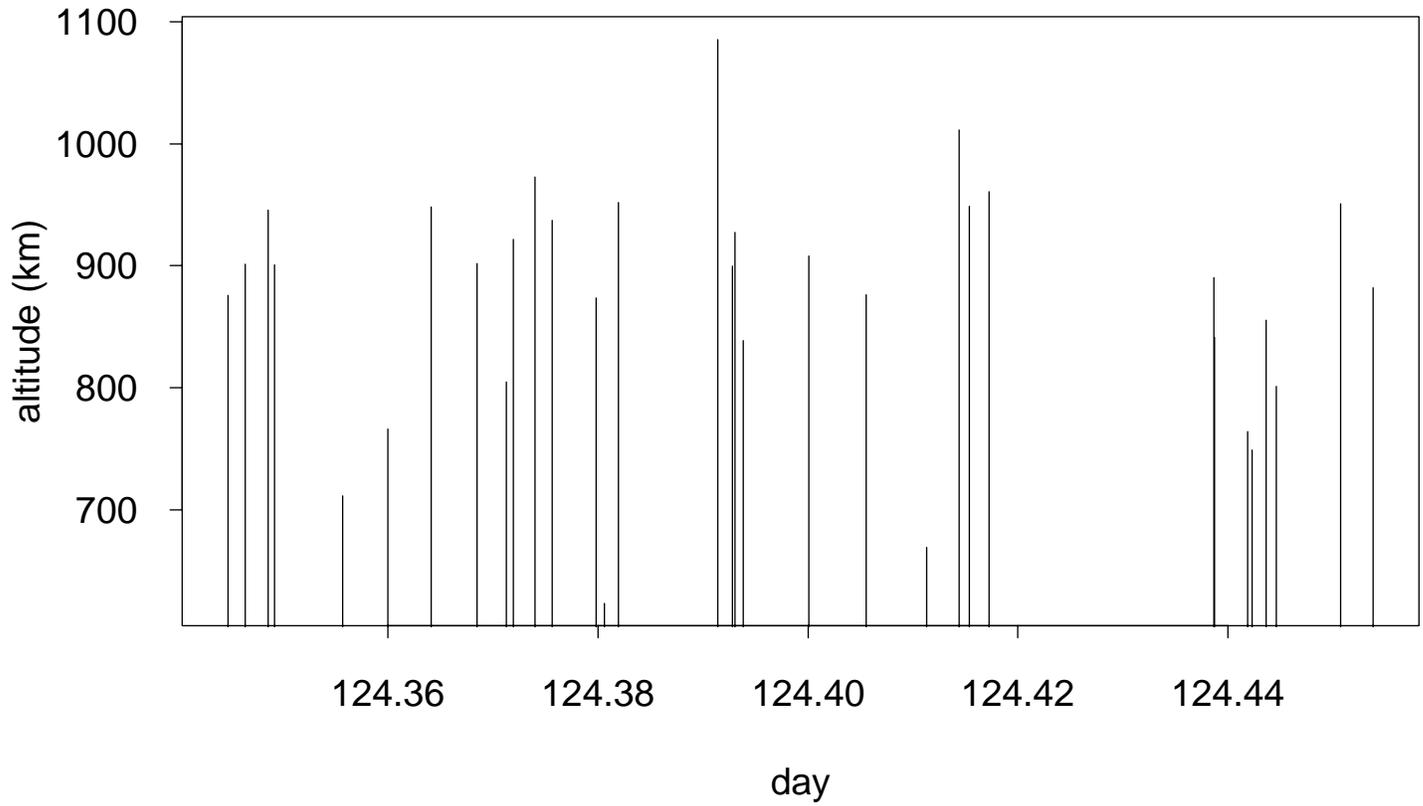
Average of periodograms of interval sequences



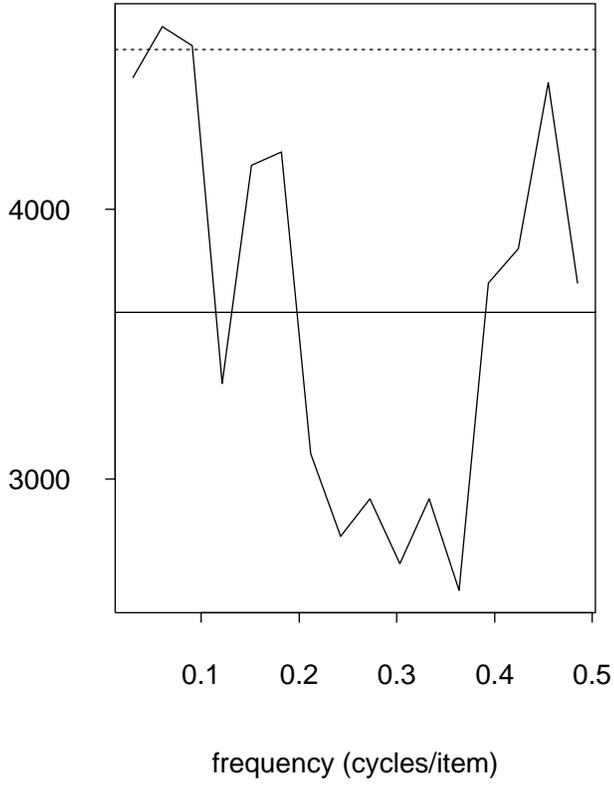
Sqrt autointensity estimate



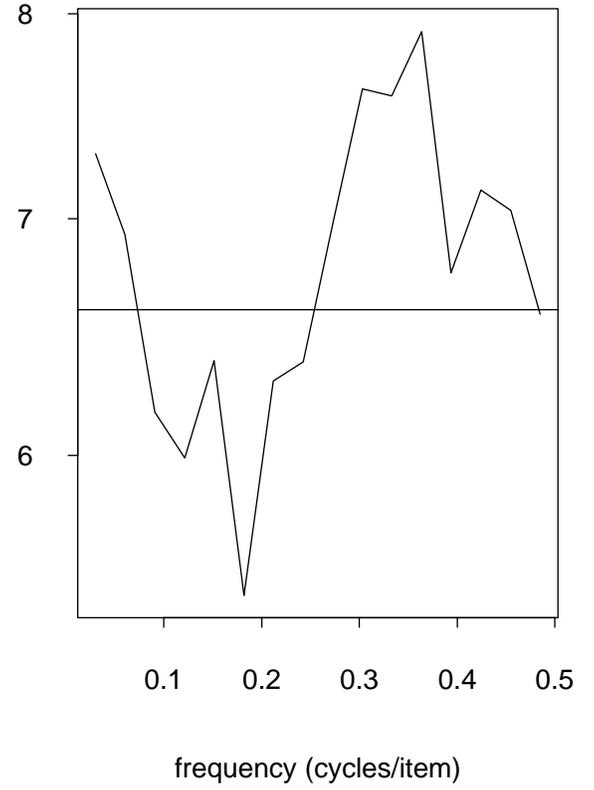
Debris pieces passing through field of view



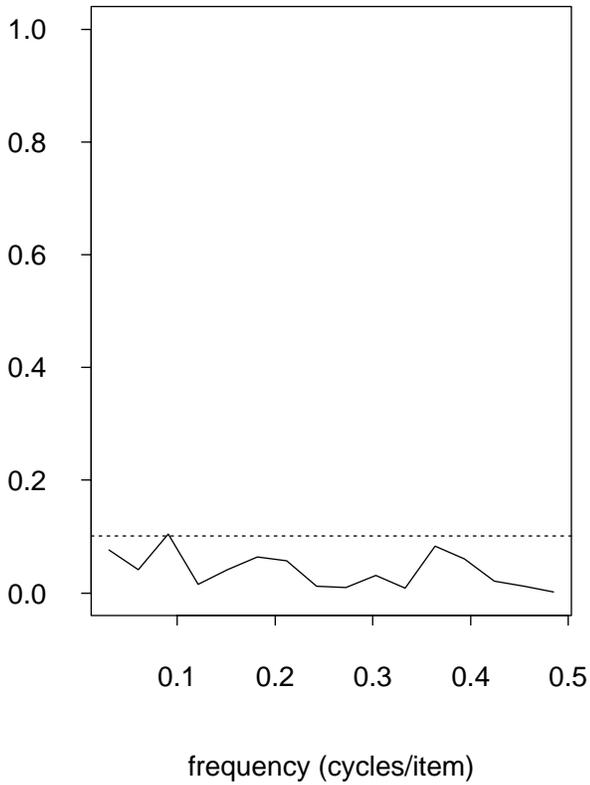
Average of periodograms of altitudes



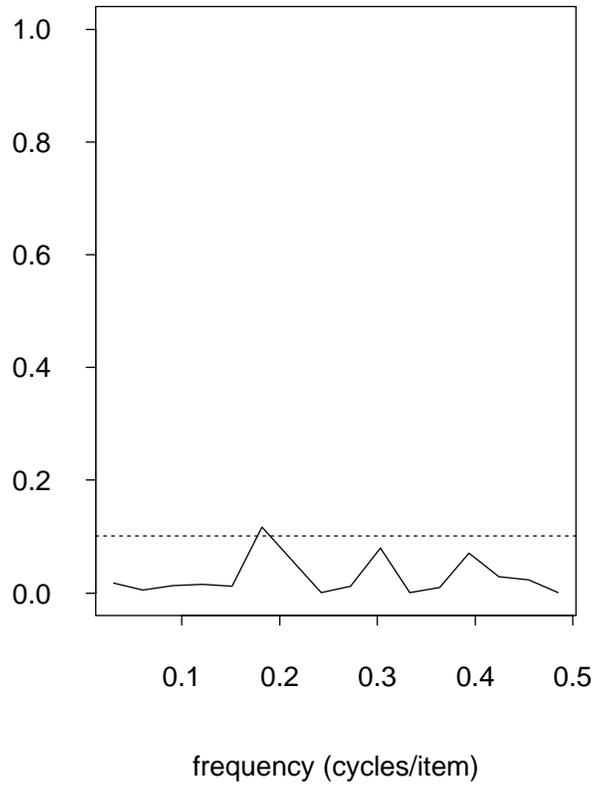
Average of periodograms of sizes



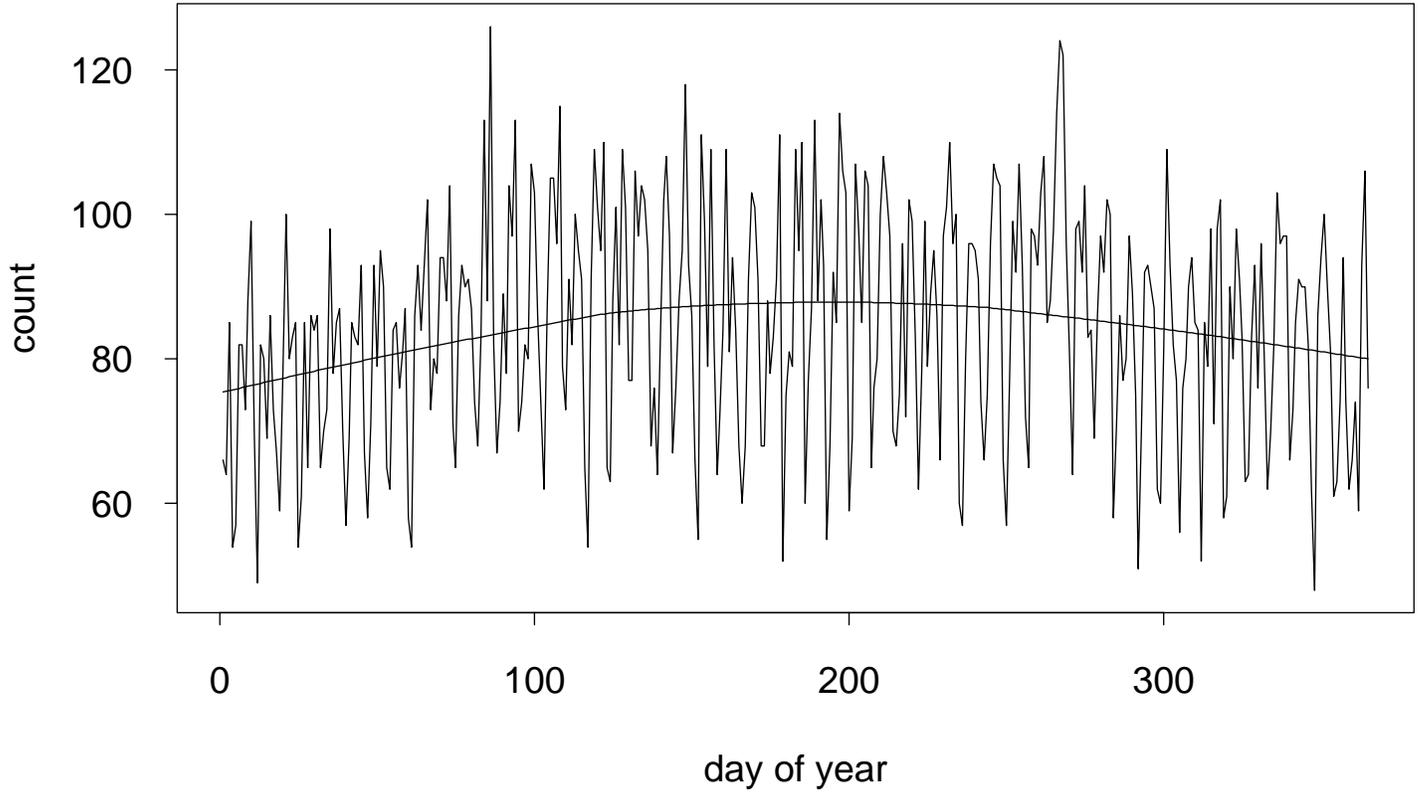
Coherence intervals and altitudes



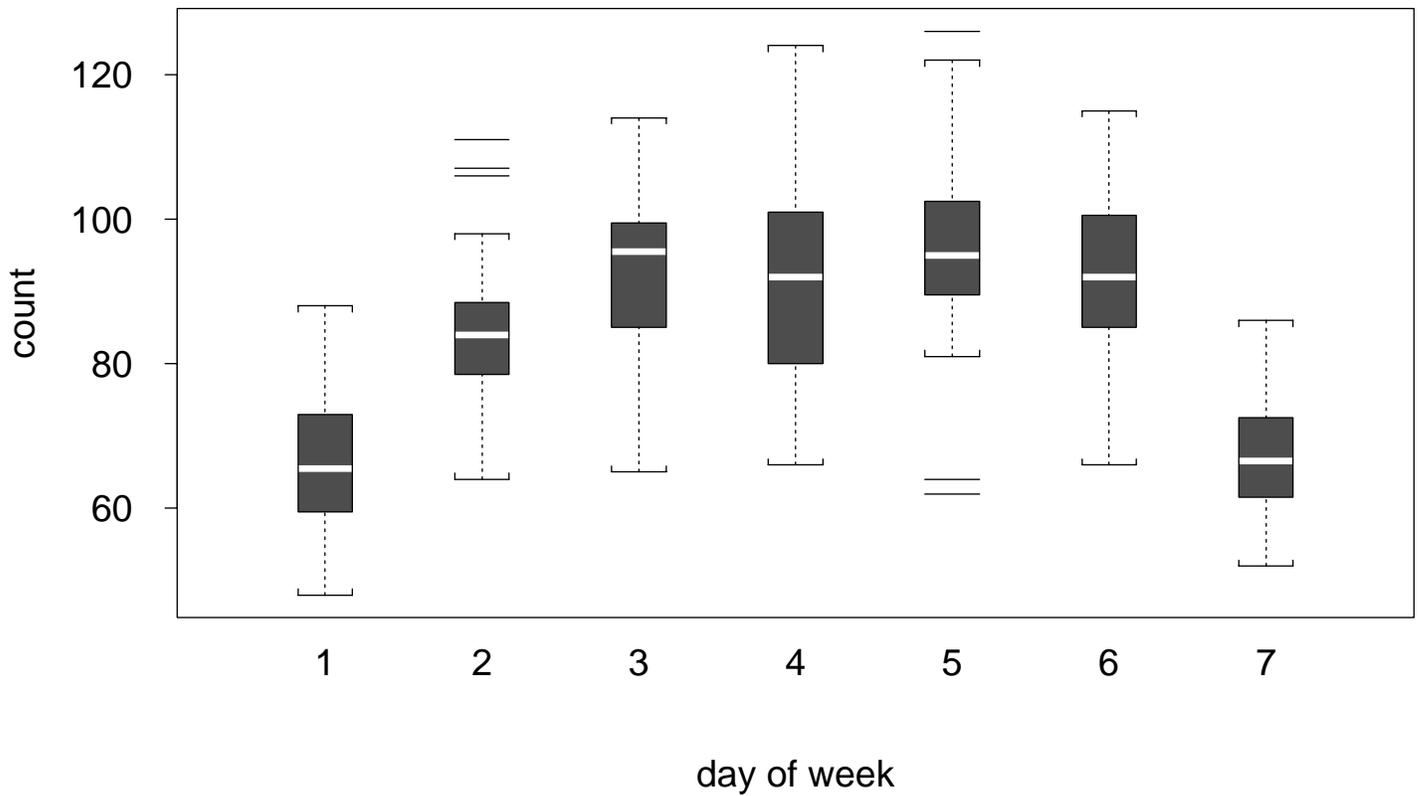
Coherence intervals and sizes



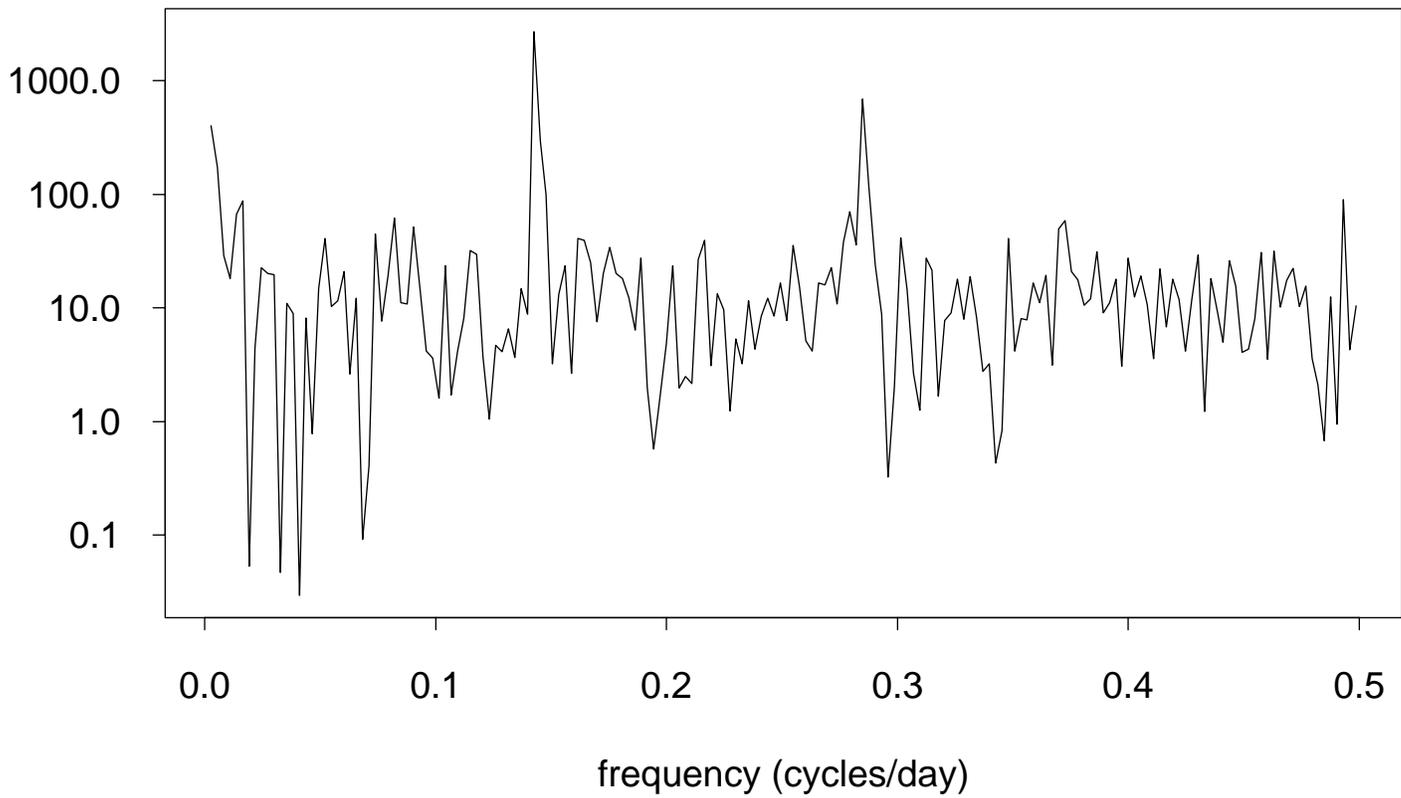
Daily births in Toronto in 1986



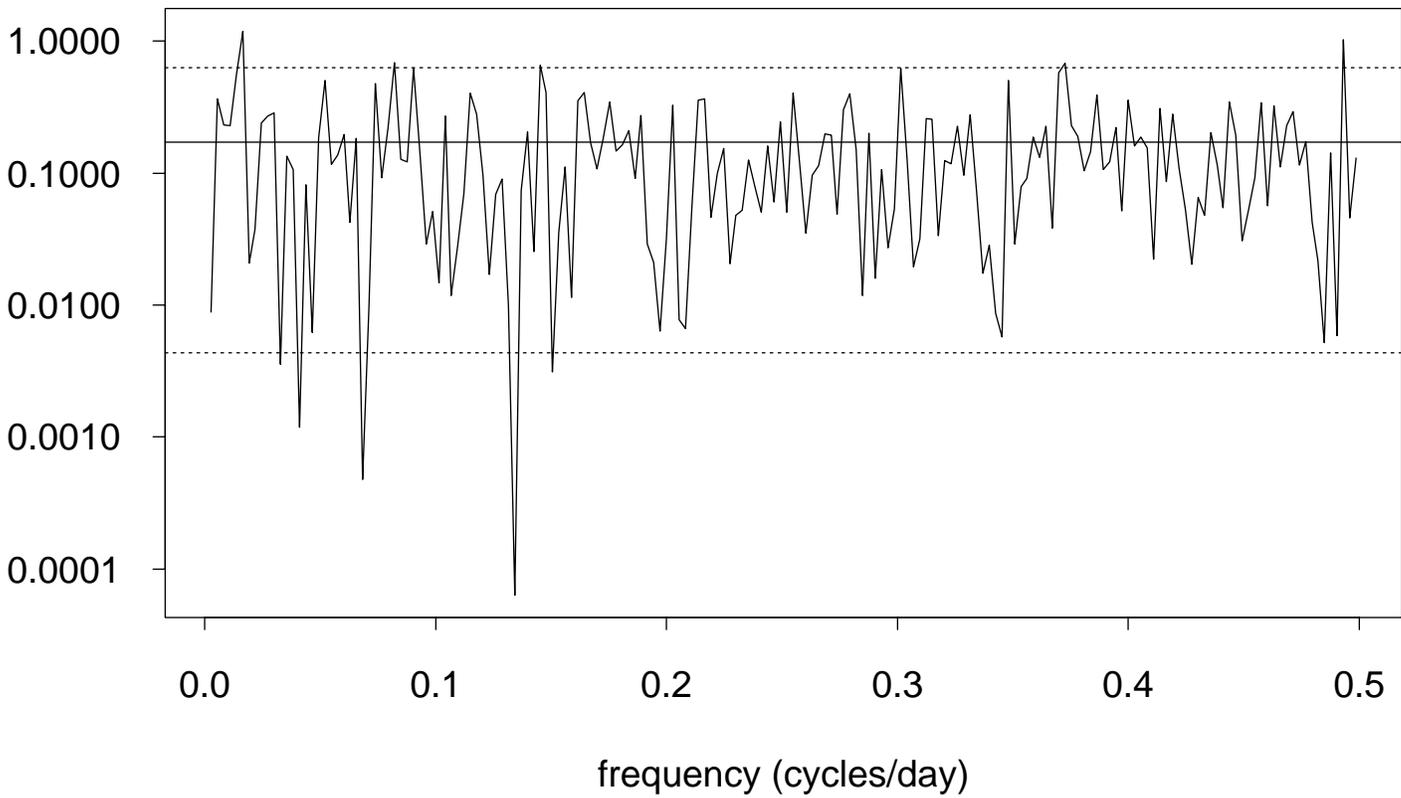
Births by day of week

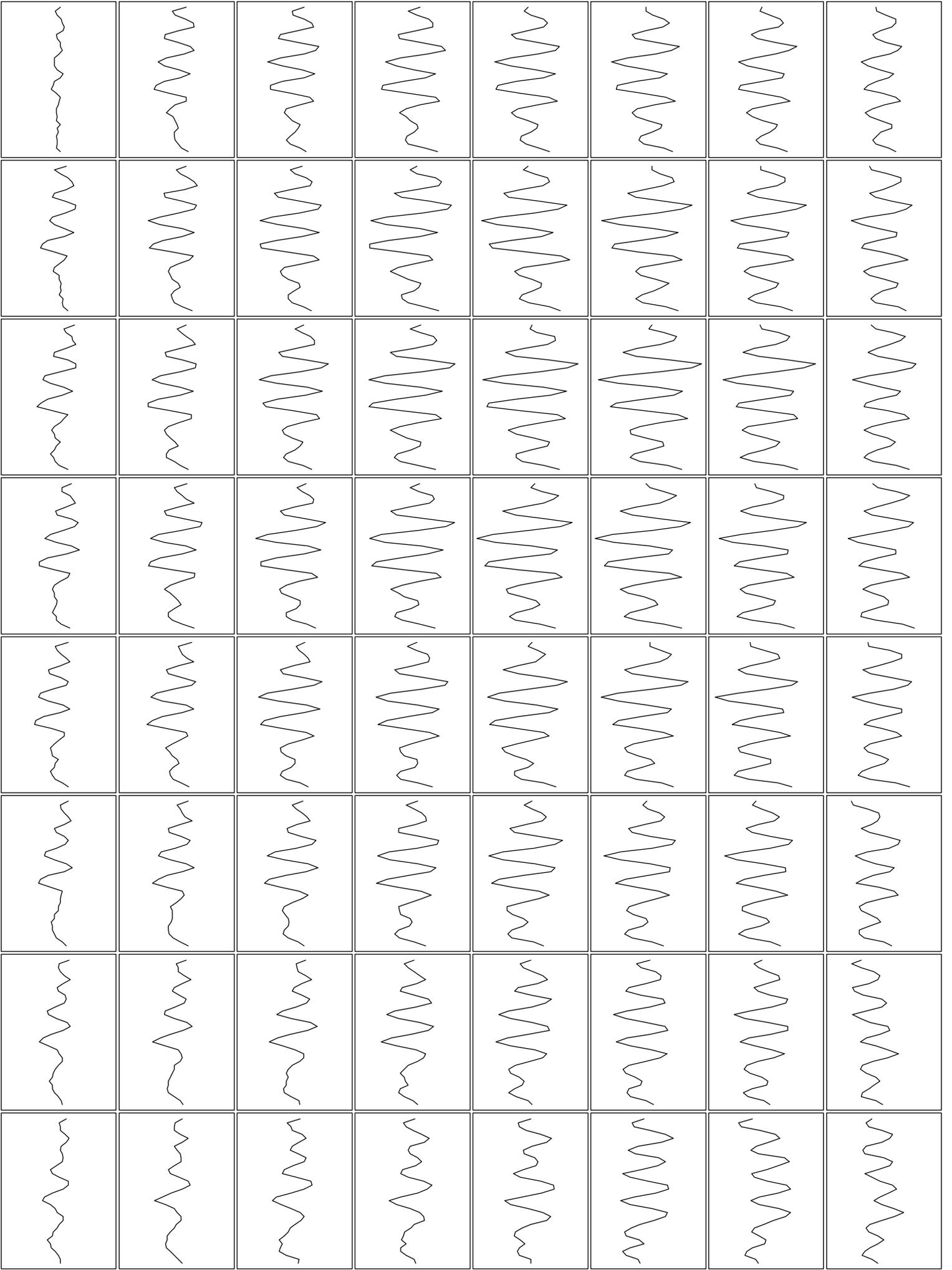


Periodogram of birth counts

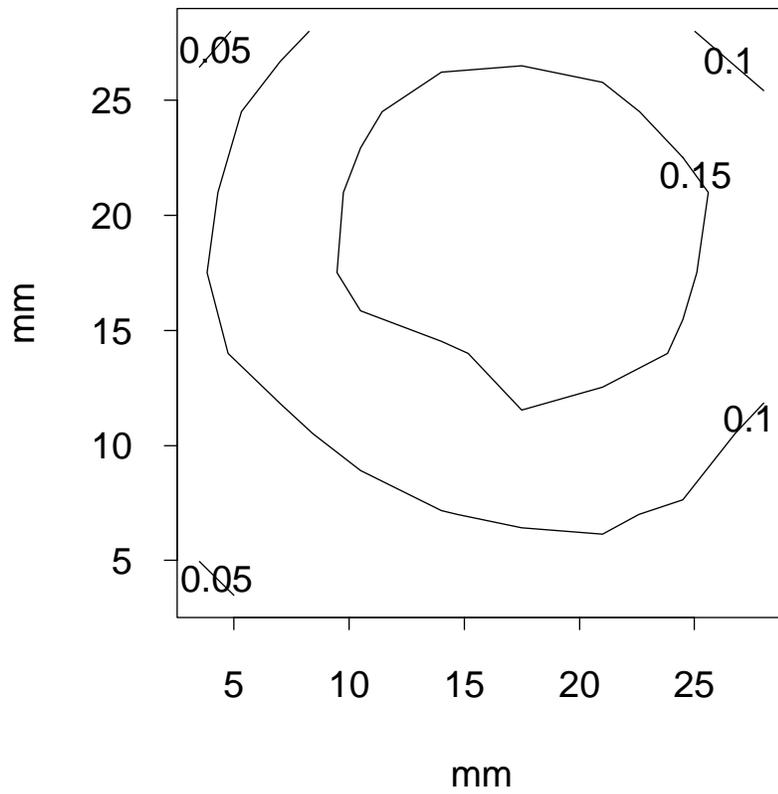
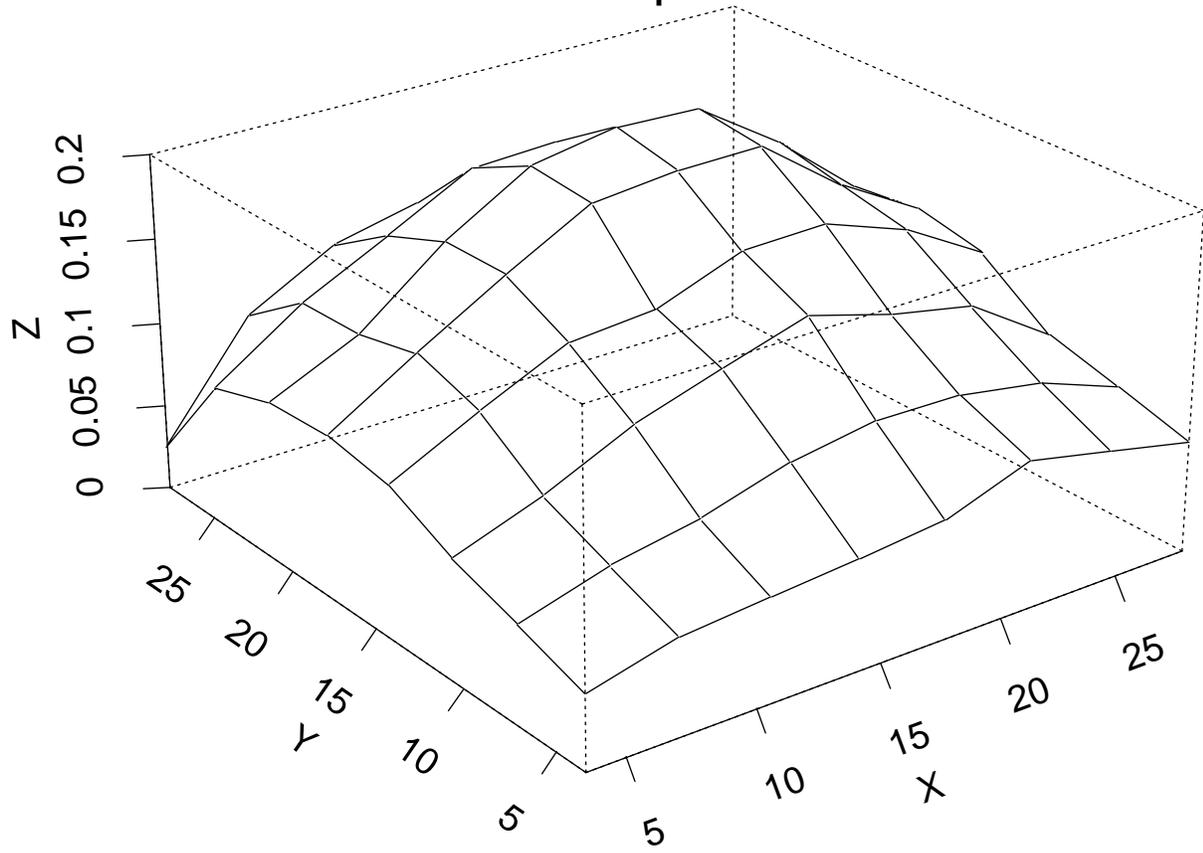


Periodogram of residuals

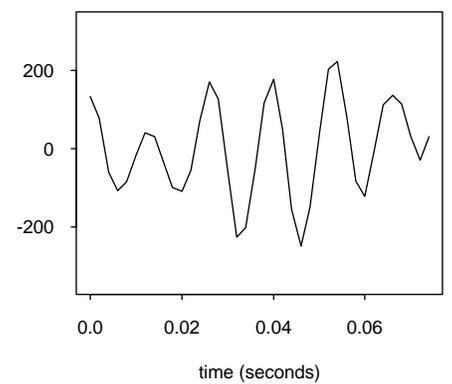
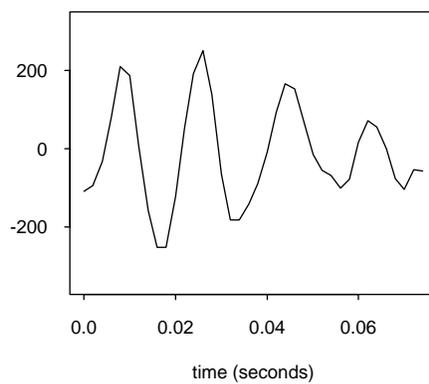
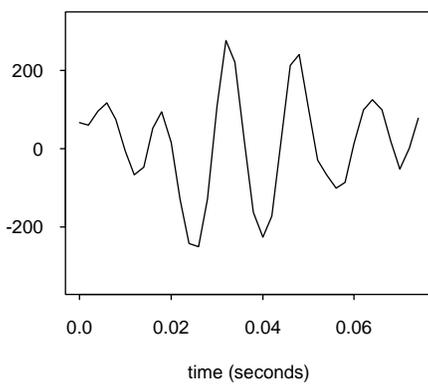
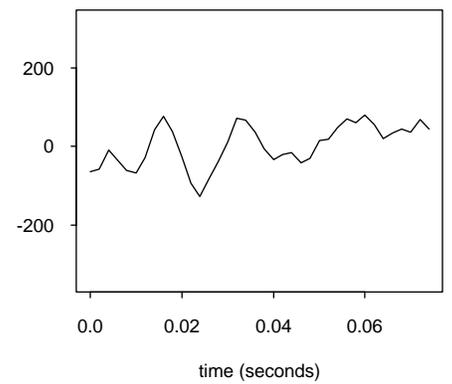
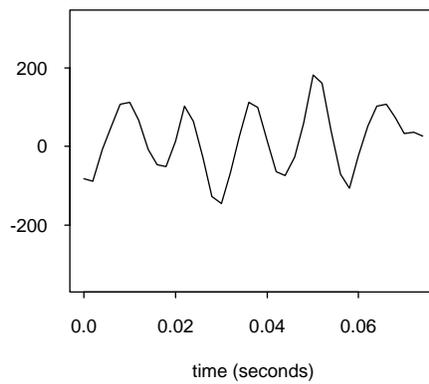
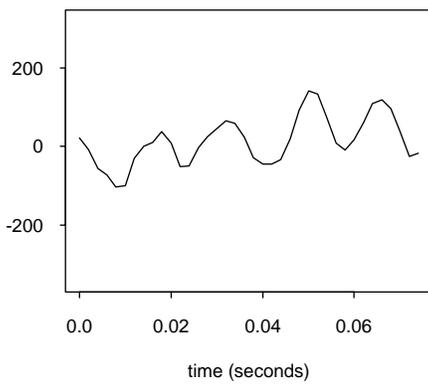
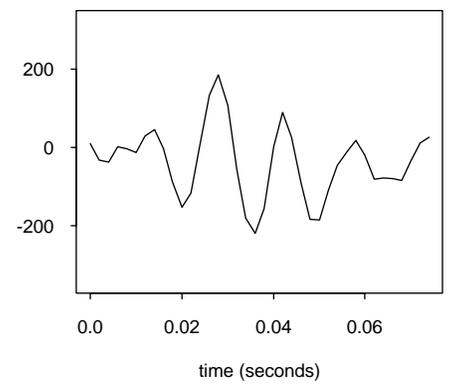
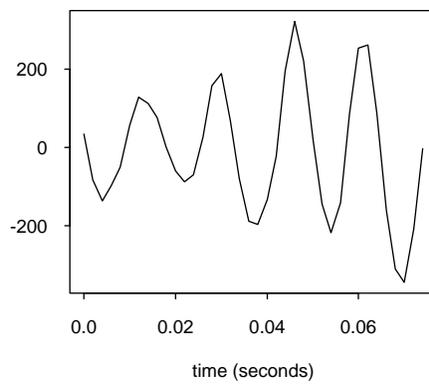
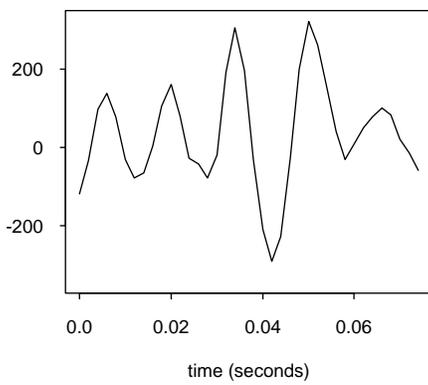
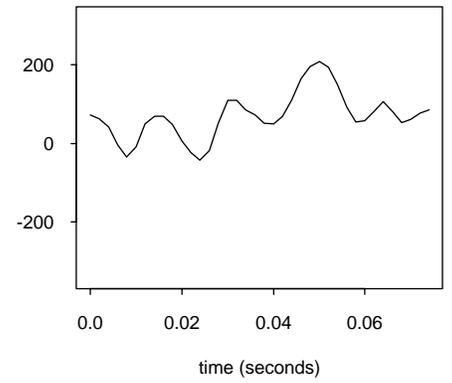
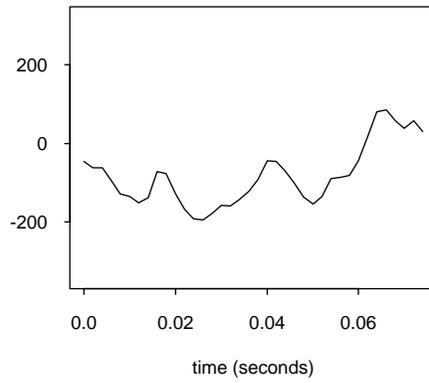
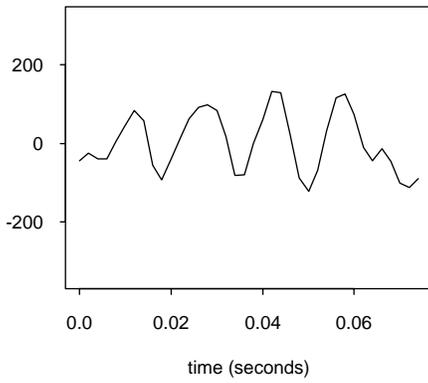


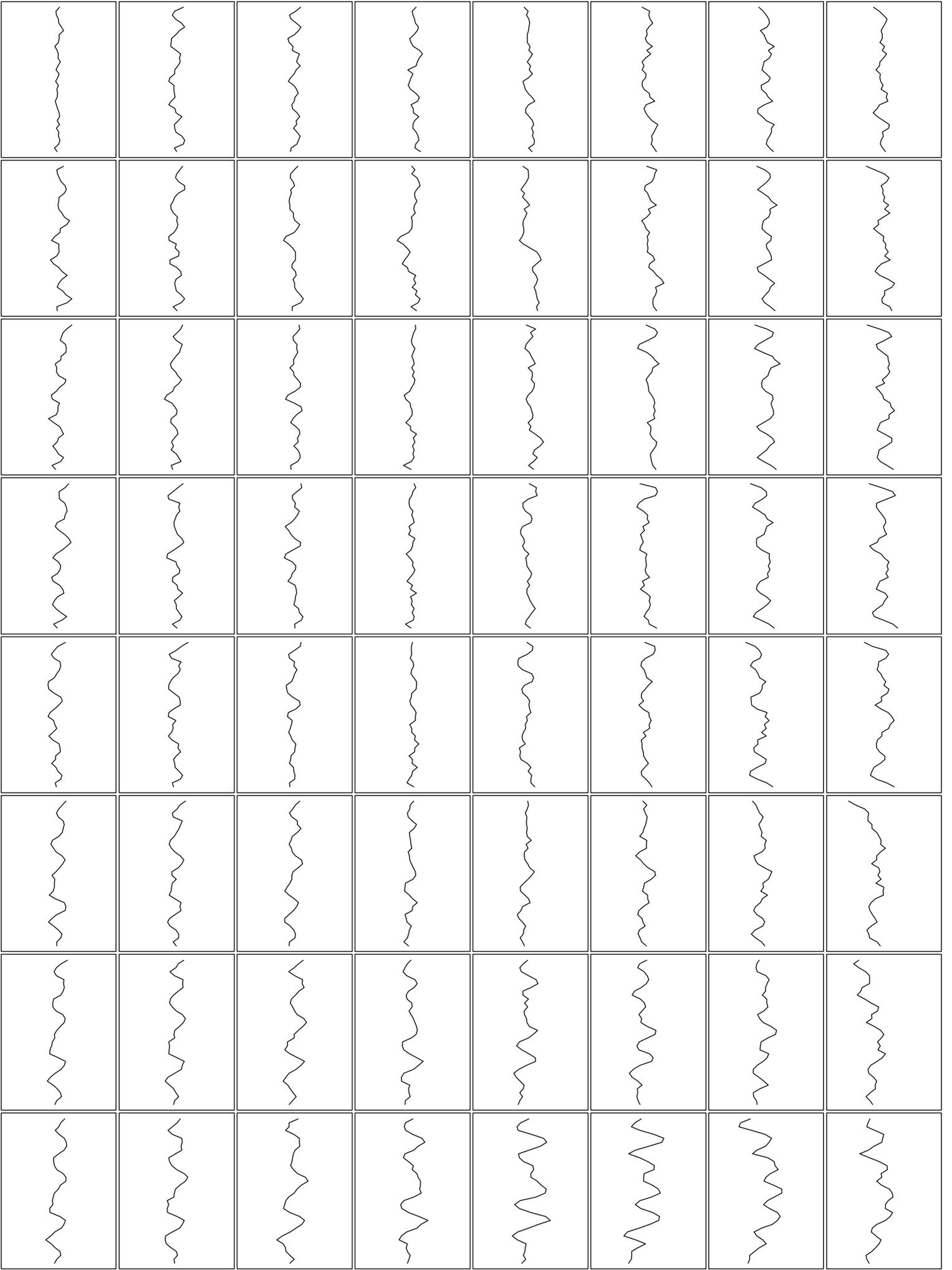


Estimated spatial function

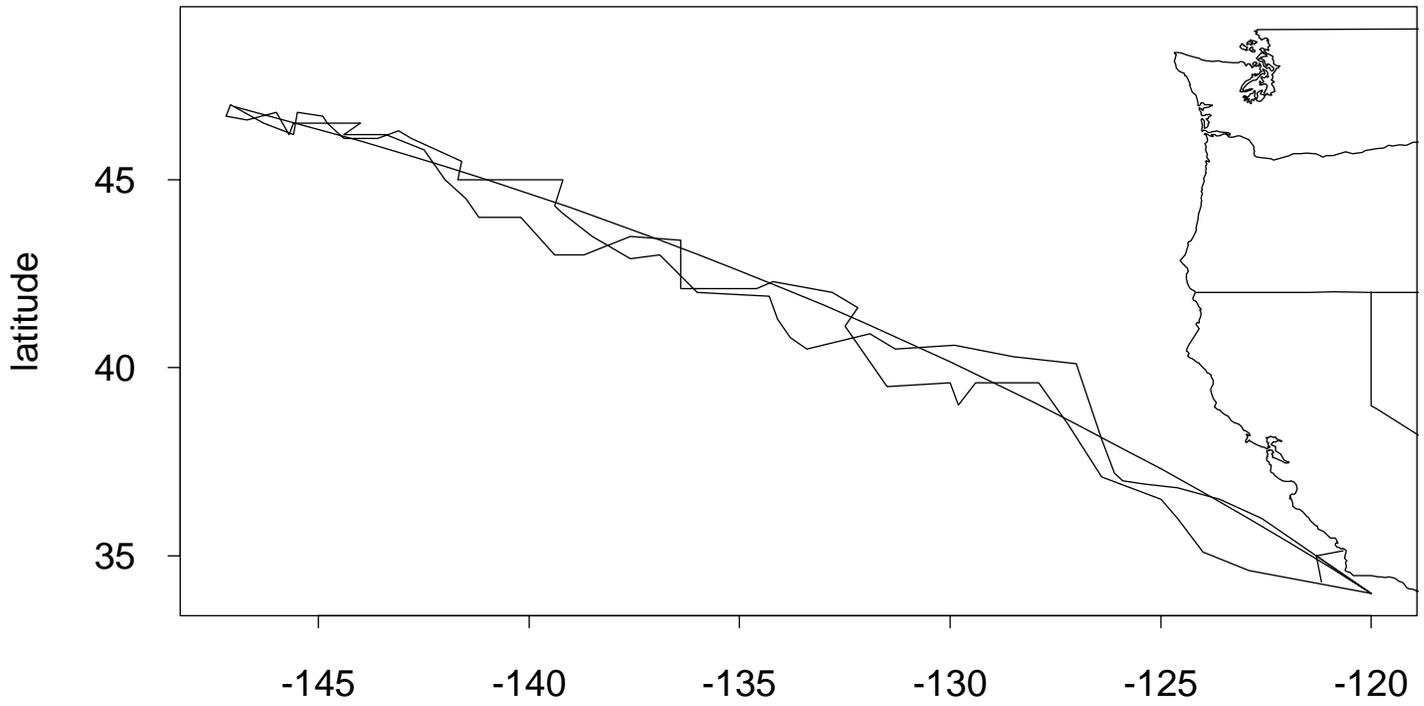


Estimated replicate time functions



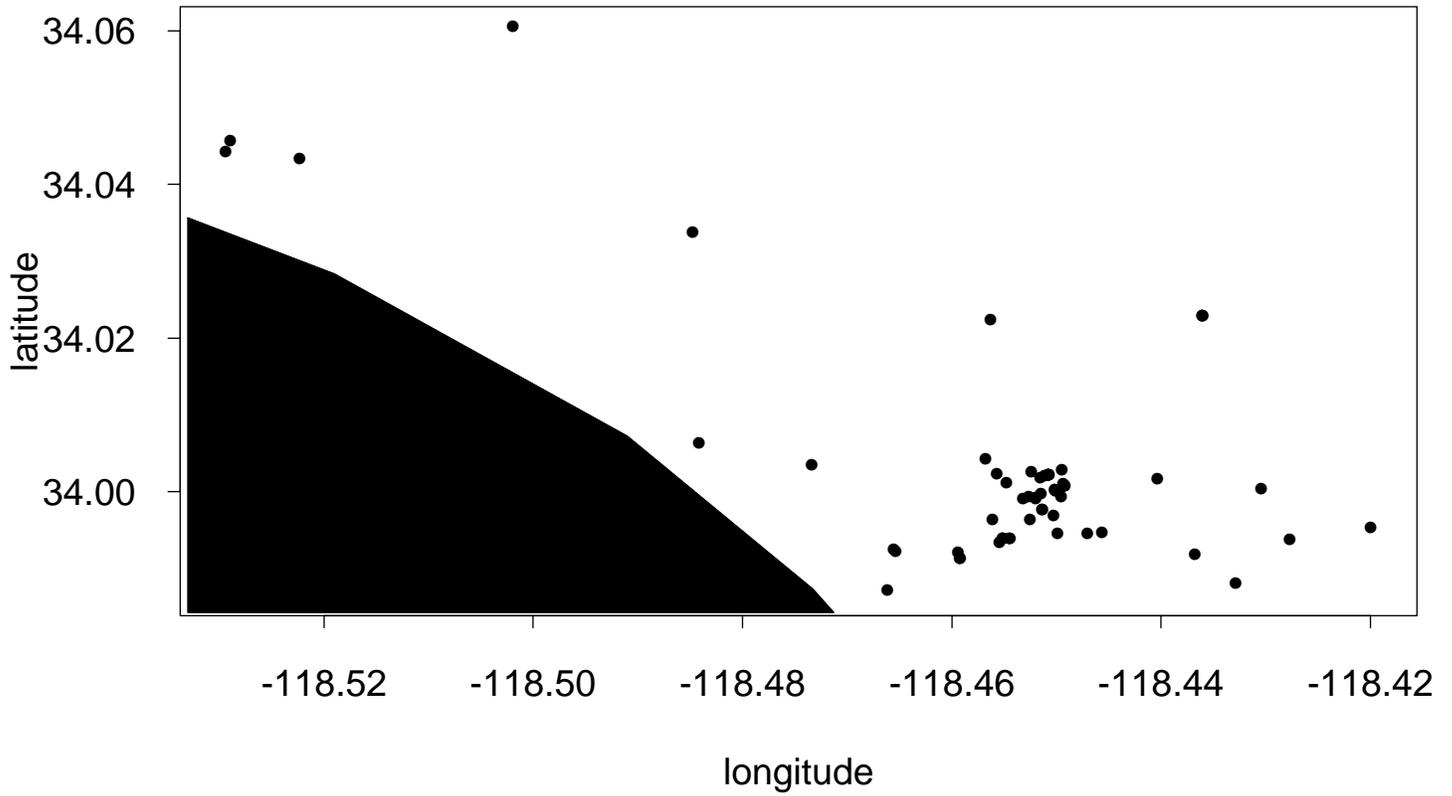


An elephant seal's migration

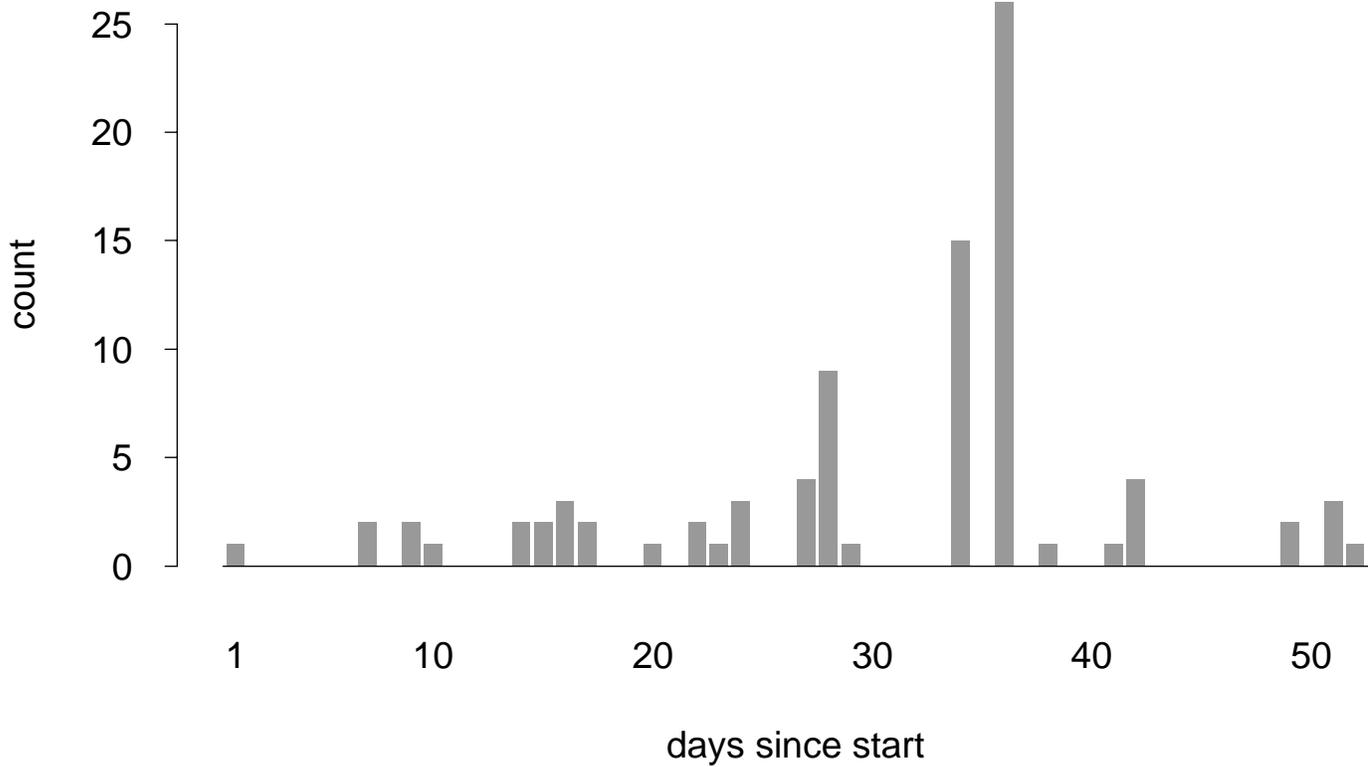


longitude
A great circle route is superposed

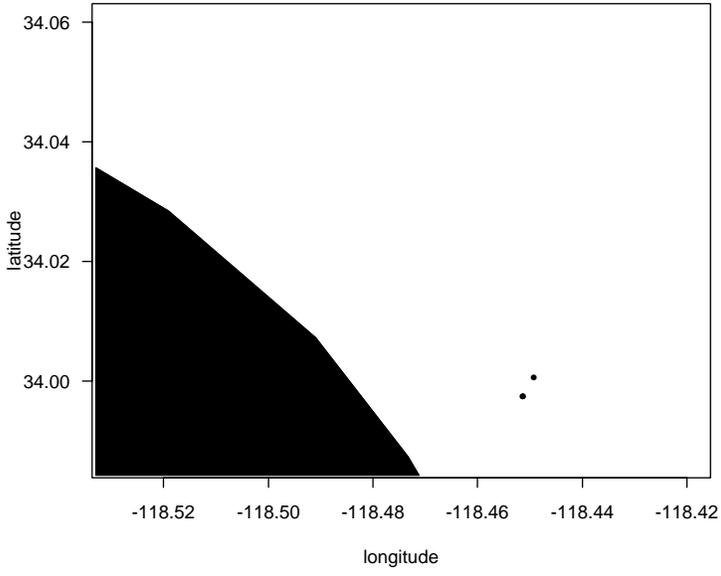
1975 medfly captures in Los Angeles, Ca.



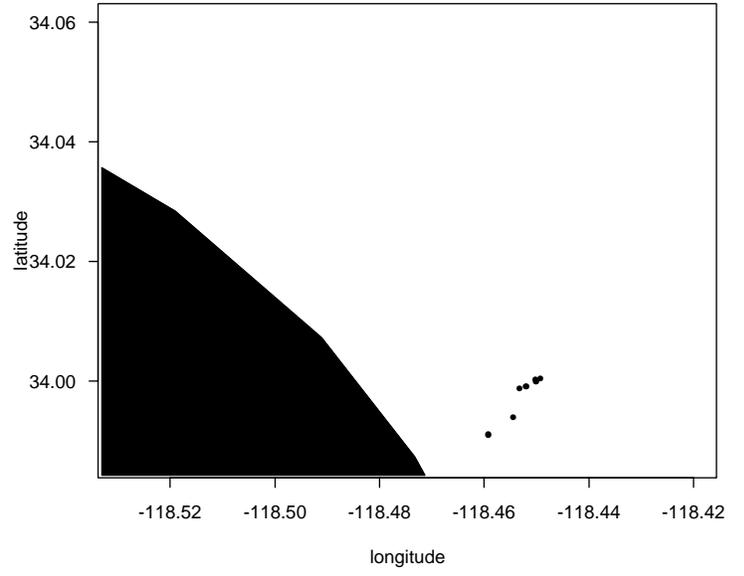
Daily medfly cases



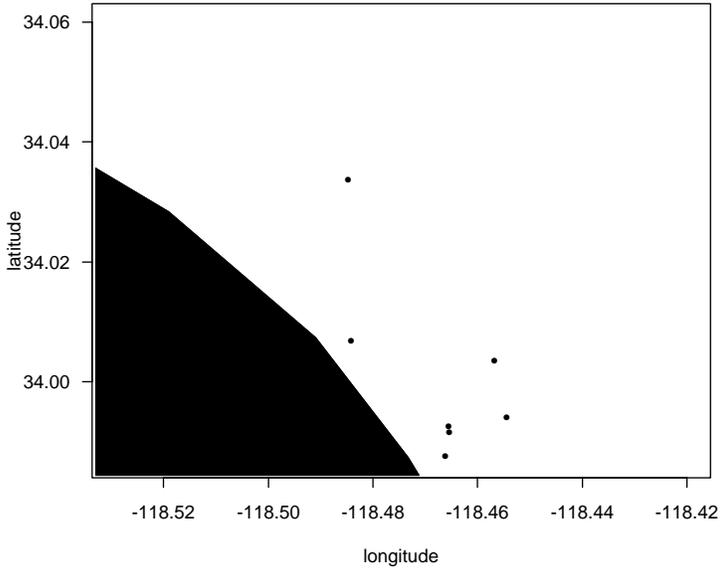
period 1



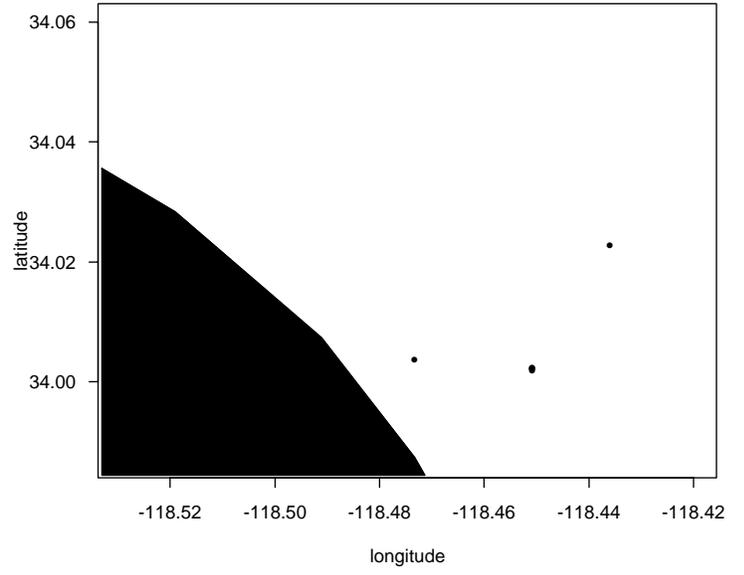
period 2



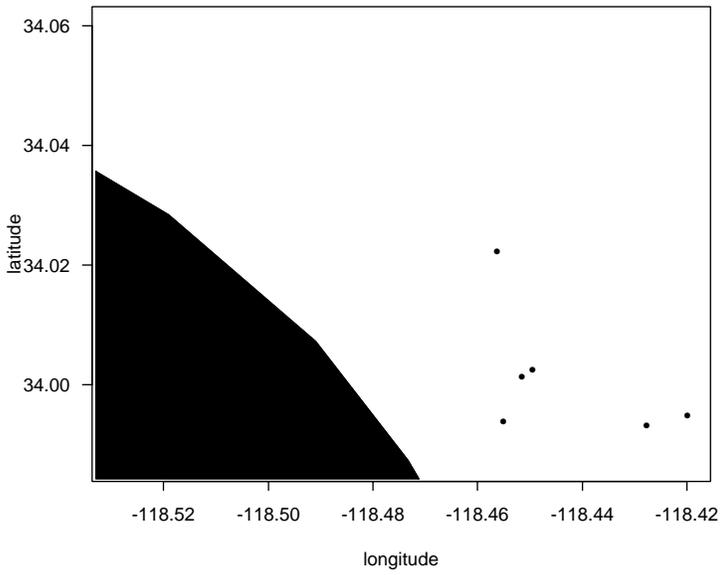
period 3



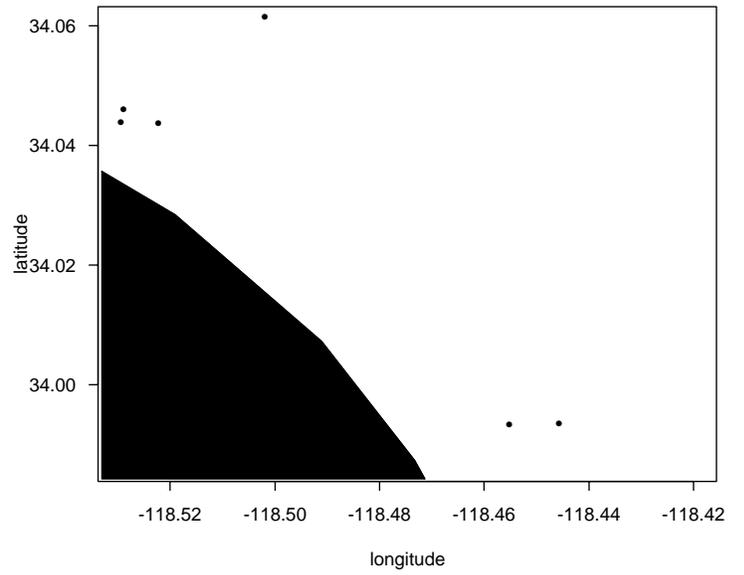
period 4



period 5



period 6



Distance from original site

