Employing stochastic differential equations to model wildlife motion

David R. Brillinger, Haiganoush K. Preisler, Alan A. Ager, John G. Kie and Brent S. Stewart

Abstract. The concern is with the properties of stochastic differential equations (SDEs) describing the motion of particles in 3 dimensional space, on the sphere or in the plane. There is consideration of the case where the drift function comes from a potential function. There is study of SDEs whose parameters are periodic in time. These are useful for incorporating circadian rhythm in the behavior. The cases of a seal in a frozen lake in Alaska, an elephant seal migrating a great distance in the Pacific Ocean and of a group of "free-ranging" elk in a reserve in Oregon are referred to. For the elk nonparametric estimates of the drift and variance terms of an SDE model are discussed and evaluated and the fit of the model assessed. One issue is how to include explanatories, beyond location and time, in the model. A number of questions motivated by the wildlife motion concerning diffusion processes of the type considered are posed at the end of the paper.

Keywords: Circadian rhythm, Diffusion model, Elk, Elephant seal, Nonparametric regression, Potential function, Ringed-seal, Stochastic differential equation, Vector field.

Mathematical subject classification: 60J60, 62G08, 62M10, 70F99.

1   Introduction

The concern is the use of vector-valued stochastic differential equations (SDEs) to describe the motion of wildlife, with the particular cases of seals and elk focused on. To illustrate the character of the data Figure 1 presents the 3D trajectory of a ringed-seal in a frozen lake in Alaska, while Figure 2 shows the path of an elephant seal migrating out into and then back from the North Pacific Ocean, and lastly Figure 3 shows an elk moving around in a reserve in Oregon. The trajectories of the animals are sampled about every 30 seconds, every 24 hours and every 2 hours respectively.
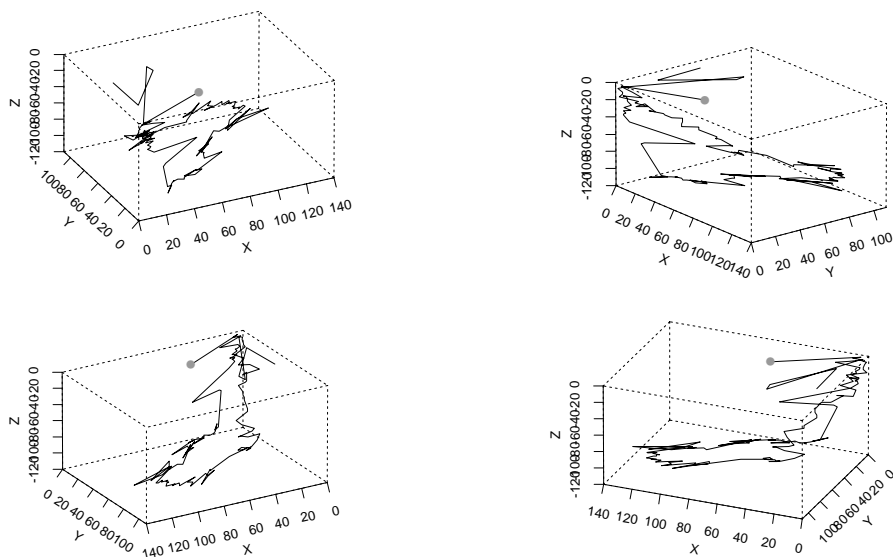
Figure 1: The trajectory of a ringed-seal diving into a frozen lake dot as seen from 4 viewpoints. The seal was released at the surface point indicated by the dot.

In the case of the seal in Alaska the lake was frozen and the concern was how did the seal find its way to air holes. The dot in the figure indicates the ice hole into which the animal was released. The seal is seen to swim to the bottom fairly directly, to swim around there for a while and then to gradually come back to the original hole. In the case of the elephant seals one wonders how they navigate? Where do they go? Is their speed approximately constant? Figure 2 indicates that sometimes they follow close to great circle routes. In the case of the elk, the reserve managers were concerned with questions like: How to allocate forage? Is change taking place? What are the effects of traffic? What is the sequence of habitat use? One can conjecture what things are important to the animals: The location of cover? The type of forage? The presence of roads? ...

An important aspect of the use of the SDEs to model such paths is that the drift and variance terms are meant to include phenomena such as: attraction, repulsion, barriers, time of day, ... There are interesting statistical questions: How to include explanatory variables? How to assess the fit of the model? How to predict future motion? ... A number of interesting questions specifically addressed to probabilists are listed in the last section of the paper.
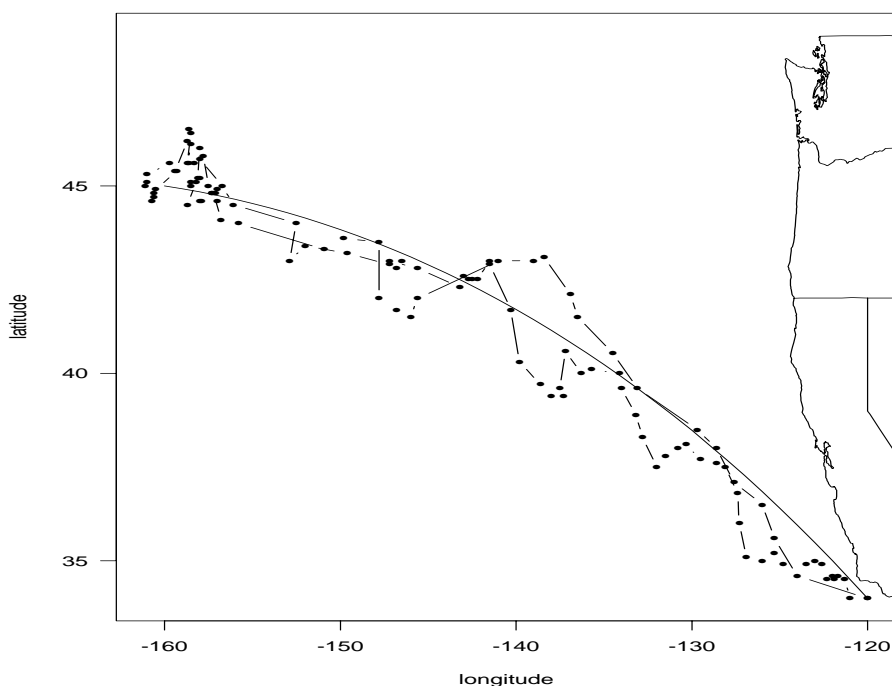
Figure 2: Points along the trajectory of an elephant seal. The curved line is a great circle route presented for comparison. The seal starts its journey from an island of Santa Barbara, California.

The paper begins with a description of deterministic and stochastic methods for describing the paths followed by particles under the influence of a field. Section 3 provides two specific models. Section 4 presents some details of the statistical methods employed. In Section 5 the experiment in which the elk data were collected is described in some detail. Section 6 presents the results obtained. Section 7 is Discussion and Summary. The final section poses some specific questions for diffusion process specialists.

References presenting models for animal movement include: Dunn and Gipson [12], Dunn and Brisbin [11], Preisler and Akers [23], White and Garrott [33], Brillinger and Stewart [7], Turchin [32]. The references White and Garrott [33], Turchin [32]) set down Fokker-Planck deterministic differential equations (DDEs) for density functions describing the expected pattern of space use. Likelihoods may be set down using conditional densities and may be used to make inferences. This deterministic approach is to be contrasted with that in the present
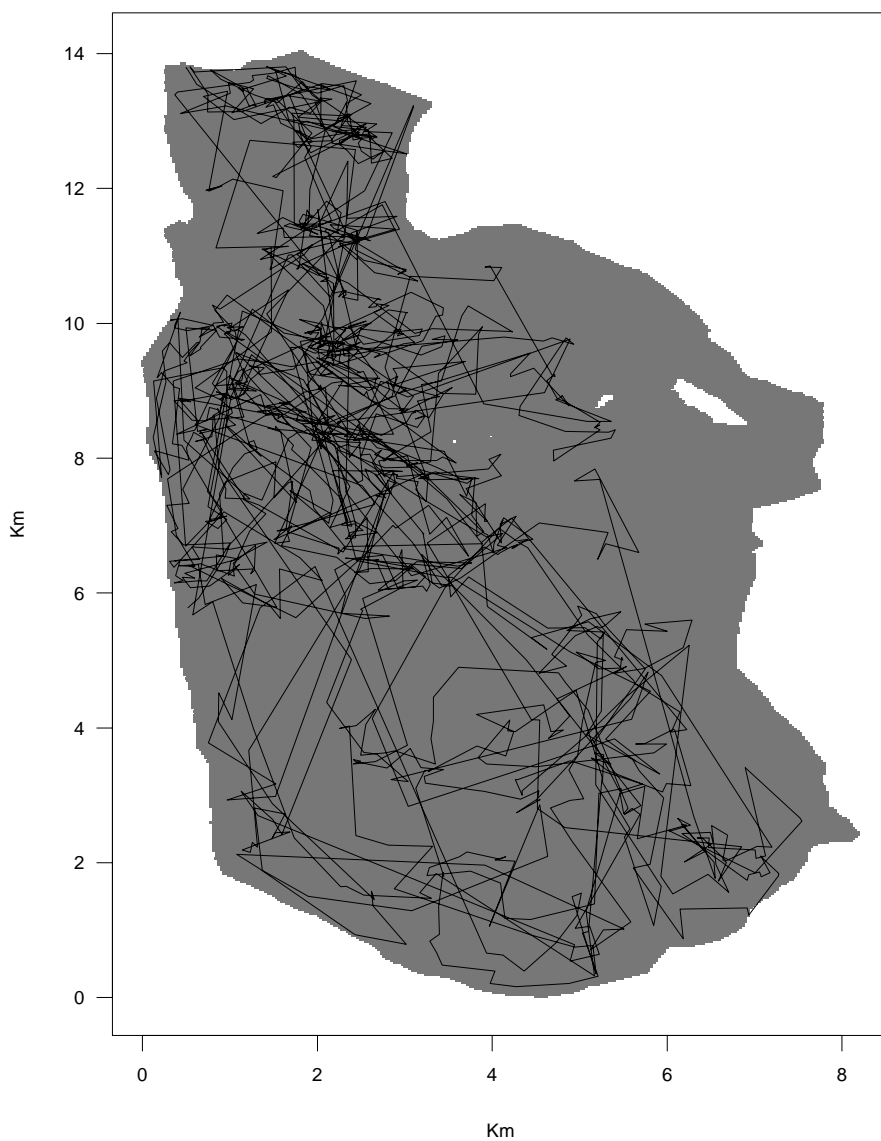
Figure 3: Points along the trajectory of one of the elk moving within the Starkey Experimental Forest.

and other papers where stochastic equations are set down describing the individual paths. Fokker-Planck equations of desired order may be derived from the SDE models. The SDE approach is appropriate for leading directly to residuals, simulation and likelihood.

2   Approaches to the description of moving particles

The analytic formulation of the motion of particles is a traditional problem of physics and applied mathematics. Classical approaches have been developed offering properties of solutions of the equations of motion and interpretations of the parameters involved. This subject matter is useful for motivating the results of the present work. First consider the deterministic approach.

2.1   Deterministic case

Motion in Newtonian dynamics may be described by a potential function, $H(\mathbf{r}, t)$, see Nelson [21]. Here $\mathbf{r} = (x, y)^\tau$ is location and t is time. The equations of motion take the form

$$
\begin{aligned}
d\mathbf{r}(t) &= \mathbf{v}(t)\, dt \\
d\mathbf{v}(t) &= -\beta \mathbf{v}(t)\, dt - \beta \nabla H(\mathbf{r}(t), t)\, dt
\end{aligned}
\tag{2.1}
$$

with $\mathbf{r}(t)$ the particle's location at time t, $\mathbf{v}(t)$ the particle's velocity and $-\beta \nabla H$ the external force field acting on the particle, β being the coefficient of friction. Here $\nabla = (\partial/\partial x, \partial/\partial y)^\tau$ is the gradient operator. The function H is seen to control the particle's direction and velocity. For example $H(\mathbf{r}) = |\mathbf{r} - \mathbf{a}|^2$ corresponds to motion with a point of attraction at $\mathbf{a}$ and $H(\mathbf{r}) = 1/|\mathbf{r} - \mathbf{a}|^2$ corresponds to motion with a point of repulsion at $\mathbf{a}$.

In the case that the relaxation time β^{-1} is small (friction is high), the equations are approximately

$$
d\mathbf{r}(t) = -\nabla H(\mathbf{r}(t), t)\, dt
\tag{2.2}
$$

and the velocity, $\mathbf{v}(t)$, is no longer involved directly, see Nelson [21].

There has been considerable mathematical development of this material e.g., Goldstein [14]. An interesting question given a force field, $\mathbf{F}$, is whether there exists a real function H, such that $\mathbf{F} = \nabla H$? When it does exist, the field is called *conservative*, see Stewart (1991) [31]. This question is addressed for the Starkey elk in Brillinger et al. [6].

2.2   Stochastic Case

A pertinent probabilistic concept for dynamic motion of a particle is a stochastic differential equation (SDE), e.g., see Nelson [21], Karlin and Taylor [17], Bhattacharya and Waymire [3]. Such equations often lead to Markov processes and take the following form

$$d\mathbf{r}(t) = \boldsymbol{\mu}(\mathbf{r}(t), t)dt + \boldsymbol{\Sigma}(\mathbf{r}(t), t)d\mathbf{B}(t) \tag{2.3}$$

Here $\mathbf{r}$, $\boldsymbol{\mu}$, $\mathbf{B}$ are vectors while $\boldsymbol{\Sigma}$ is a matrix. The drift $\boldsymbol{\mu}$ may be interpreted as a velocity field and an example of an estimate will be provided later in the paper. $\boldsymbol{\Sigma}$ is the variance or diffusion parameter while $\mathbf{B}$ is a random function such as a Brownian or Levy process. The parameters and the Brownian process control the direction and speed of motion.

The parameters have interpretations provided by

$$E\{d\mathbf{r}(t)|H_t\} = \boldsymbol{\mu}(\mathbf{r}(t), t)\, dt$$

and

$$var\{d\mathbf{r}(t)|H_t\} = \boldsymbol{\Sigma}(\mathbf{r}(t), t)\, dt$$

with dt is small and H_t representing the time history of the process, $\{\mathbf{r}(u), u \leq t\}$. The driving process $\mathbf{B}$ leads to variability around deterministic motion. This process might correspond to explanatories omitted from the equations. The vector $\boldsymbol{\mu}(\mathbf{r}(t), t)$ is seen to represent the instantaneous velocity of the particle at time t and position $\mathbf{r}$. Since the process is Markov, these conditional moments depend only on the preceding position, $\mathbf{r}(t)$.

A variety of properties are known concerning solutions of SDEs, for example when H does not depend on t and $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\tau = \sigma_0^2\mathbf{I}$, there is often an invariant density

$$\pi(\mathbf{r}) = c\, exp\{-2H(\mathbf{r})/\sigma_0^2\} \tag{2.4}$$

representing the longrun density of locations that the process visits; see Bhattacharya and Waymire [3]. Thus in such a stationary case, by analyzing the paths, population densities may be estimated.

The situation of a particle being affected by the force field of a potential function may be visualized by picturing a ball rolling around in the interior of a perspective plot of the potential function. In the stochastic case there is a convenient approximation for viewing the situation. First consider a one-dimensional process $dx = \mu(x, t)\, dt + \sigma(x, t)dB(t)$. Suppose that at time t the

particle is at location $x(t) = x$. Then for the particle's location at time $t + dt$ take

$$x(t + dt) = x \pm \sigma(x, t)\sqrt{dt} \quad \text{with prob} \quad \frac{1}{2} \pm \frac{\mu(x, t)}{2\sigma(x, t)}\sqrt{dt}$$

see Kloeden and Platen [19]; Prohorov and Rozanov [26]. As $dt \to 0$ the paths of this process approach those of the SDE with parameters μ, σ. In the bivariate case one generates two such processes. Examples are given in Brillinger et al. [6].

A random potential/environment may be a useful model. It might be used to describe seals chasing fish, elk attracted or repelled by other elk. The randomness is introduced by the motion of these other particles. For the potential function one might write

$$H(\mathbf{r}, t) = \alpha(\mathbf{r}, t) \sum_1^J |\mathbf{r} - \mathbf{X}_j(t)|^2$$

with the $\mathbf{X}_j(t)$ locations of moving attractors/repellors. This could be the source of the $d\mathbf{B}(t)$ term in (2.3).

The components H_x, H_y of the gradient ∇H correspond to the components of $\boldsymbol{\mu}$ of (2.3). One advantage of the potential function approach is that independent potential functions from a variety of sources add as in

$$\sum_l H_l(\mathbf{r}, t)$$

There may be points, lines or regions of attraction or repulsion and barriers to be included. The barriers can represent actual physical objects (e.g., fences). The process $\mathbf{B}$ includes impulses due to the presence of objects such as trees, mounds, dips and natural variability corresponding to individual animal behavior. The fitted SDEs may be used to produce estimates of other parameters, (e.g. expected speed and distances to road), to predict spatial and temporal patterns of animal distribution and habitat preferences, to simulate trajectories and to study the directionality of the movement for example.

3   Some specific models

3.1   Ornstein–Uhlenbeck

A particular case of an SDE that has been employed in describing animal motion is the mean-reverting *Ornstein–Uhlenbeck* (O–U) process; see Dunn and

Gipson [12]; Dunn and Brisbin [11]. Here

$$\mu(\mathbf{r}, t) = \mathbf{A}(\mathbf{a} - \mathbf{r})$$

and

$$\Sigma(\mathbf{r}, t) = \Sigma$$

while the mean is $\mathbf{a}$. The O–U process becomes the *random walk* when $\mathbf{A} = \mathbf{0}$, i.e., when the drift term, $\mu(\mathbf{r}, t)$, is $\mathbf{0}$.

If $\mathbf{A}$ is symmetric and positive definite, the corresponding potential function is

$$H(\mathbf{r}, t) = (\mathbf{a} - \mathbf{r})^\tau \mathbf{A}(\mathbf{a} - \mathbf{r})/2$$

and the particle is wandering but being pulled towards the location $\mathbf{a}$. The invariant distribution is multivariate normal, $N(\mathbf{a}, \Psi)$, with

$$\Psi = \int_0^\infty e^{-\mathbf{A}u} \Sigma \Sigma^\tau e^{-\mathbf{A}u} du;$$

see p. 597 in Bhattacharya and Waymire [3]. If $\Sigma\Sigma^\tau = \sigma_0^2\mathbf{I}$, then $\Psi = \sigma_0^2\mathbf{A}^{-1}/2$.

3.2   Diffusion on a sphere

In the case of the elephant seals one can consider the case of a diffusion process with drift on the sphere, see Brillinger [4]. With $\phi(t)$ denoting latitude and $\theta(t)$ denoting colatitude the equations of the process are

$$d\theta_t = \sigma dU_t + \left(\frac{\sigma^2}{2\tan\theta_t} - \delta\right) dt$$

$$d\phi_t = \frac{\sigma}{\sin\theta_t} dV_t$$

Here δ is the drift parameter and (U_t, V_t) denotes a bivariate Brownian motion.

4   The statistical methods used

One can consider both parametric and nonparametric methods of inference. Dunn and Gipson [12] use approximate maximum likelihood procedures to estimate the parameters $\mathbf{a}$, Ψ of the multivariate O–U process from sampled trajectories with constant sampling intervals. Dunn and Brisbane [11] give extensions of the maximum likelihood estimate to the case where observations are unequally spaced over time.

Turning to a more general situation, the Radon-Nikodyn derivative of the process (2.3) with respect to the measure of $\mathbf{r}(t) = \sigma(\mathbf{r}, t)d\mathbf{B}(t)$ is

$$\exp\left\{-\int_0^T \sigma(\mathbf{r}, t)^{-1}\mu(\mathbf{r}, t) \cdot \sigma(\mathbf{r}, t)^{-1}d\mathbf{r} + \int_0^T |\sigma(\mathbf{r}, t)^{-1}\mu(\mathbf{r}, t)|^2 d\mathbf{r}/2\right\}$$

[22], and estimates of a finite dimensional parameter may be computed by maximization. In the sampled time case the SDE (2.3) may be approximated as follows

$$(\mathbf{r}(t_{l+1}) - \mathbf{r}(t_l))/(t_{l+1} - t_l) \approx \mu(\mathbf{r}(t_l), t_l) + \Sigma(\mathbf{r}(t_l), t_l)\mathbf{Z}_l/\sqrt{t_{l+1} - t_l} \quad (4.1)$$

$l = 1, 2, \ldots$ with $t_1 < t_2 < t_3 < \cdots$ sampling times and with the $\mathbf{Z_l}$ independent bivariate standard normals. This follows from (2.3) directly. The validity of the approximation is investigated in Kloeden and Platen [19].

In terms of the individual components (X, Y) of $\mathbf{r}$ one has from (4.1)

$$\frac{\Delta X}{\Delta t} = \mu_x(X, Y, t) + \text{noise}$$

$$\frac{\Delta Y}{\Delta t} = \mu_y(X, Y, t) + \text{noise} \qquad (4.2)$$

If the drift function components, μ_x, μ_y, are smooth, one sees that one has a nonparametric regression estimation problem. Principal statistical tools employed in the work are smoothing methods and residual plots. The smoothing or nonparametric regression procedure employed to estimate μ_x, μ_y is the function *loess*() of Cleveland et al. [9] within *gam*() of Hastie [15]. It involves the local fitting of quadratics in the explanatories.

In the example of the elk, time of day proves important and conditional density estimates are evaluated for selected times of day. Estimates will be provided of conditional longrun population densities and of functions depending smoothly on time and location and explanatories. There are a variety of such estimates including: kernel-based, spline-based, local polynomial, Hastie and Tibshirani [16].

References to inferential methods for diffusion processes, both parametric and nonparametric, include Banon and Nguyen [1], Bertrand [2], Burgière [8], Dohnal [10], Genot-Catalot et al. [13], Sorensen [29], Prakasa Rao [22].

A question is how to include explanatory variables corresponding to phenomena such as: fences, roads, streams, forage, cover, terrain, time of day, times of sunrise and sunset, periods of hunting. Attraction/repulson phenomena might be described by a potential term $H(\mathbf{r}, t) = \alpha d(\mathbf{r})^\beta$ where $d(\mathbf{r})$ is the shortest distance to a feature.

5   The Elk experiment

The main study area at Starkey Experimental Forest and Range is 7,762 ha large and located in the Blue Mountains of northeastern Oregon, Rowland et al. [27]. The forest was enclosed with a 2.4-m tall, net-wire fence in 1988 and radio-telemetry studies were begun. Each spring a number of elk, deer, and cattle are fitted with collars containing Loran-C receivers. The collars are instructed to intercept Loran-C broadcasts at regular intervals and then to relay those signals to a central receiver. Locations are then estimated from the time delays.

The study area is also managed for a variety of public uses such as recreation, hunting, forest management, cattle grazing, and other activities. The shape of the area is shown in Figure 3. The white areas in the figures correspond to two small elk-proof exclosures within the study area.

The data are spatial-temporal. The locations of $M = 53$ elk, (labelled by $m = 1, \ldots, M$, and recorded at times, $t_{mk}, k = 1, \ldots, K_m$ for the m-th animal) are given as well as various explanatory variables describing vegetation and topography. Other habitat features (e.g., distance to road, distance to water) suspected to influence elk movement, are also available. The locations are written as a column vector $\mathbf{r}_{mk} = (X_m(t_{mk}), Y_m(t_{mk}))^\tau$, corresponding to the UTM (Universal Transverse Mercator) coordinates of the k-th time measurement of the m-th elk.

The data used in the work of this paper were collected between April 7 and November 15, 1994 and involve 53 female elk. Observations were omitted for 30 days during the autumn of 1994 when hunting was conducted within the project area. Preliminary analyses of these data often showed erratic movements that were not typical of the rest of the year. Also omitted were movements where the time interval exceeded 1.5 hours or was less than 0.1 hours. Velocities calculated from short time intervals are strongly influenced by telemetry error while long time intervals generate uncertainty regarding the true trajectories of elk.

Figure 3 showed the successive estimated locations for one of the elk. The trajectories plotted are a sequence of straight-line segments and therefore are jagged. This discreteness results from the fact that location estimates are only available once every 0.1-4 hours. Figure 5, to follow, will show the estimated distribution of all 53 animals at four different times of the day.
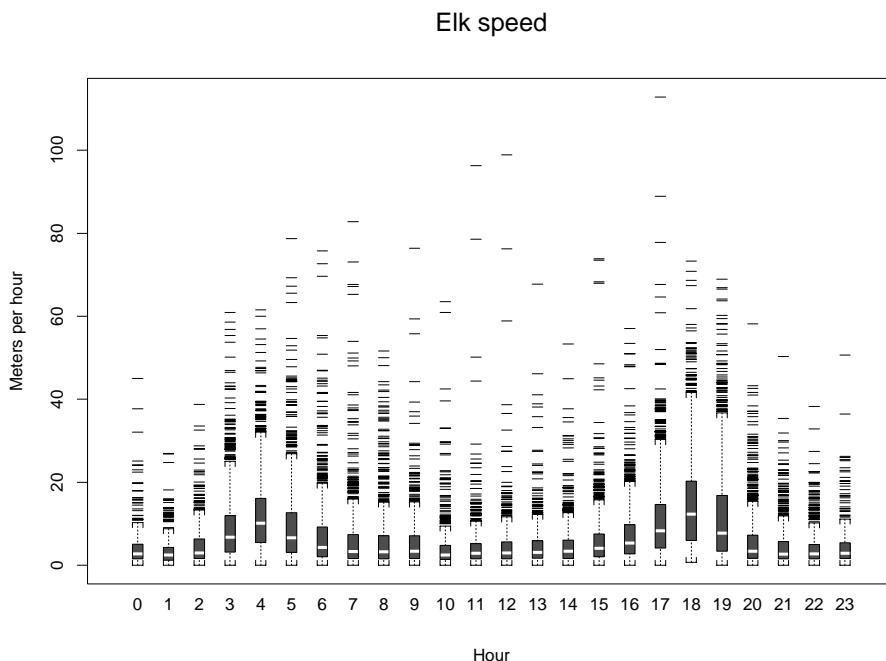
Figure 4: Estimated speed for all 53 elk as a function of time of day.

6   Results of analyses

6.1   Some descriptive statistics

It is natural to imagine a dependence of an elk's movements on time of day and on location. Temporal dependence can be anticipated from the circadian rhythm (with a period of 24 hours) in animal behavior. Consider Figure 4, a parallel boxplot of elk speeds by hour of the day. The elk are seen to be more mobile around 0400 hrs and 1800 hrs and are less active at night and midday. These observations agree with previous studies of elk that have shown strong activity cycles that are characterized by dawn and dusk transitions between foraging and resting habitat. Speed was estimated by the distance between two successive locations divided by the difference of the observation times.

   Turning to the spatial aspect, Figure 5 gives a density estimate of location obtained by using a kernel estimate. The estimate is computed for two hour long time periods centered at the times 0600, 1200, 1800, 2400, i.e. equi-sampling the day. The darker pixels correspond to greater density. There are hot spots of high density and cold spots of avoidance. As computed, these densities provide
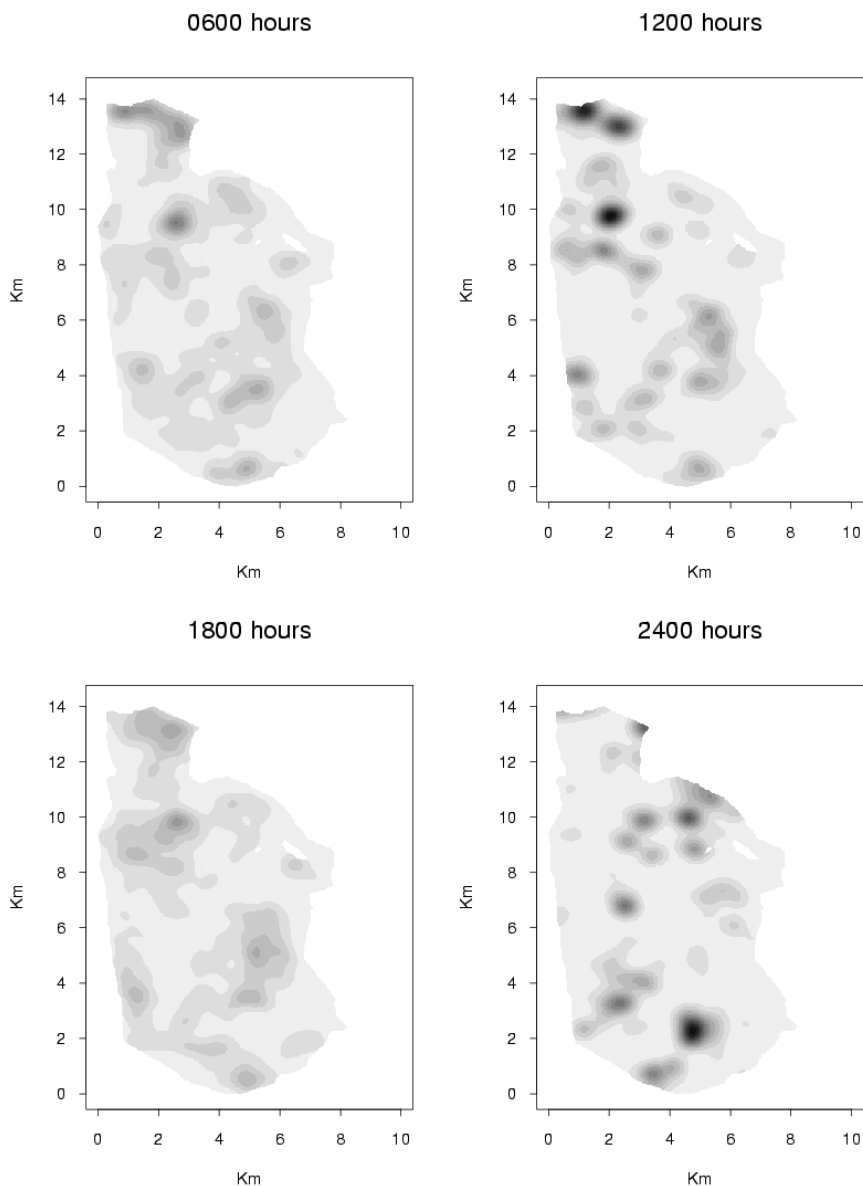
Figure 5: Density estimate based on the data of all 53 elk at 4 times of day. Darker values correspond to higher density values. The contours are equispaced from 0 and the four displays are on the same scale. The highest corresponds to 5 elk/km^2.

the long run distribution of elk locations for the indicated time intervals. From Figure 4 the most active periods appear centered at times 0400 and 1800, while the least active are around 1100 and 2300. These times correspond approximately to the equispaced ones used in Figure 5.

From a comparison of the four separate images, the time effect does not appear to be simply proportional rather there is an interaction of time and location. This possibility may be investigated more formally. A simple dot plot, see Brillinger et al. [6], of the locations the animals visit suggests that the sample of elk were well-distributed about Starkey during the study interval.

6.2   Modelling μ

Following the model (4.2) the problem of learning about the vector field μ may be seen as one of nonparametric regression analysis. Estimates of the functions $\mu_x(\cdot)$, $\mu_y(\cdot)$ of (4.2) were calculated via the function $gam()$ of Hastie [15] making use of the function lo() of Cleveland et al. [9]. Following (3.1) the weights $t_{m,k+1} - t_{m,k}$ were used for the m-th elk.

The preliminary investigations of the previous section suggested that elk movements were affected by both time of day and location. One wonders if these effects are additive. To study this possibility the model $\mu(\mathbf{r}, t) = \mathbf{g}(\langle t \rangle)$, $\mathbf{g}$ being supposed smooth, was first fit where $\langle t \rangle$ denotes the time of day at time t. Then the additive model, $\mu(\mathbf{r}, t) = \mathbf{g}(\langle t \rangle) + \mathbf{h}(\mathbf{r})$ was fit. Finally the general model, $\mu(\mathbf{r}, t) = \mathbf{i}(\mathbf{r}, \langle t \rangle)$, $\mathbf{i}$ being supposed smooth, was employed. This succession of models was employed in order to look for simplifications in the structure. The spans used in $lo(\cdot)$ were .4, .16, 064 following Hastie (1992)'s page 276 suggestion for obtaining approximately the same marginal span and to make the analyses nested. (The specific values were picked to be on the small side, yet to give stable estimates.) The resulting F-values were all 0 to the accuracy of Splus, and assumed that the F-distribution was appropriate. The degrees of freedom (DF) were those produced by $gam()$. The formula for the DFs and the accompanying F-values are motivated in Hastie (1992) by an assumption of independent gaussian errors and fixed regressors. When an analysis of variance was carried out the resulting F-values were all 0 to the accuracy of the Splus computations. The relation thus appears nonadditive. This complicates the display of the estimates. Figure 6 presents the estimated $\mu(\mathbf{r}, \langle t \rangle)$ for times of day $\langle t \rangle = 0600, 1200, 1800, 2400$ using the final i.e. nonadditive model in time of day and location. The figures are vector-field plots with the lengths of the arrows proportional to the estimated $\sqrt{\mu_x{}^2 + \mu_y{}^2}$ at the indicated locations. The angles of the vectors give the estimated direction of motion away from these

0600 hours

1200 hours
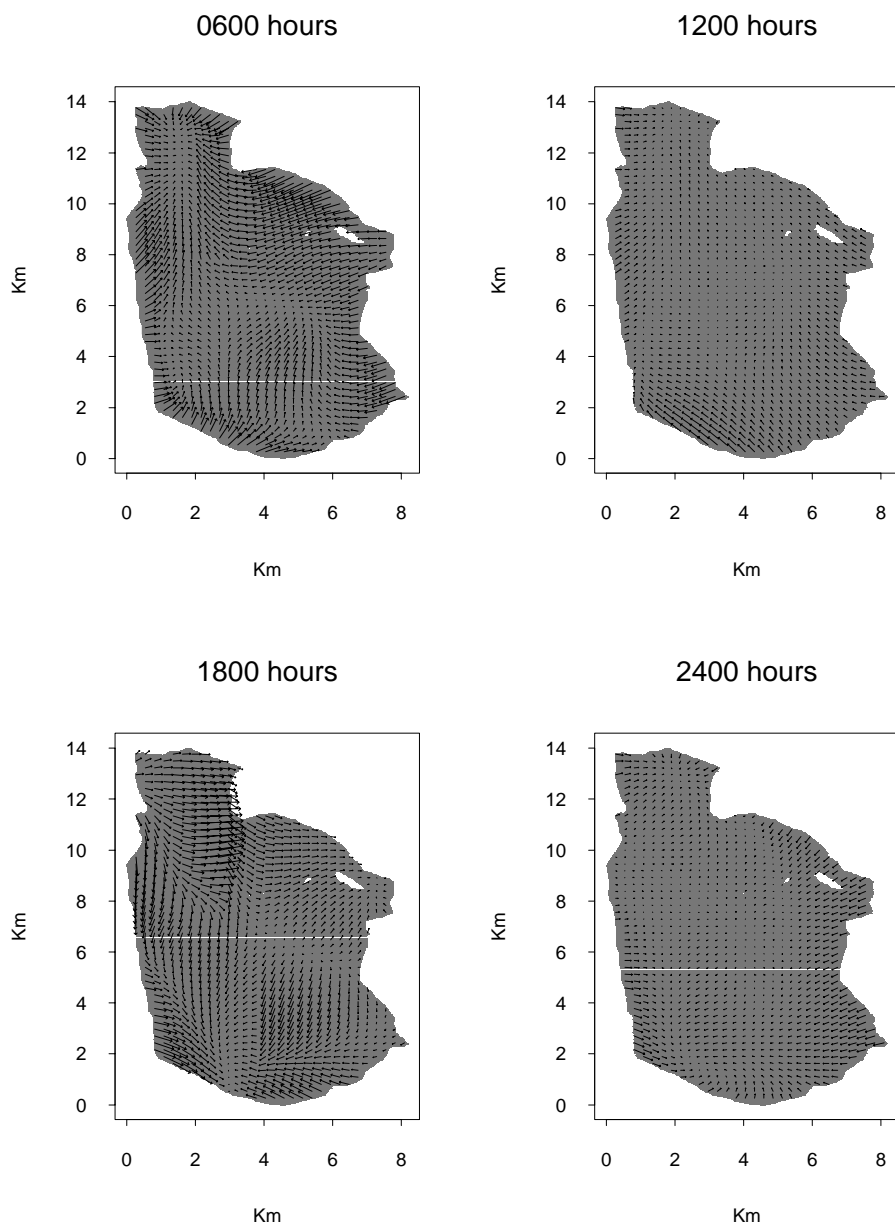
1800 hours

2400 hours

Figure 6: The estimated gradient vector field for times of day 0600, 1200, 1800, 2400 hours.
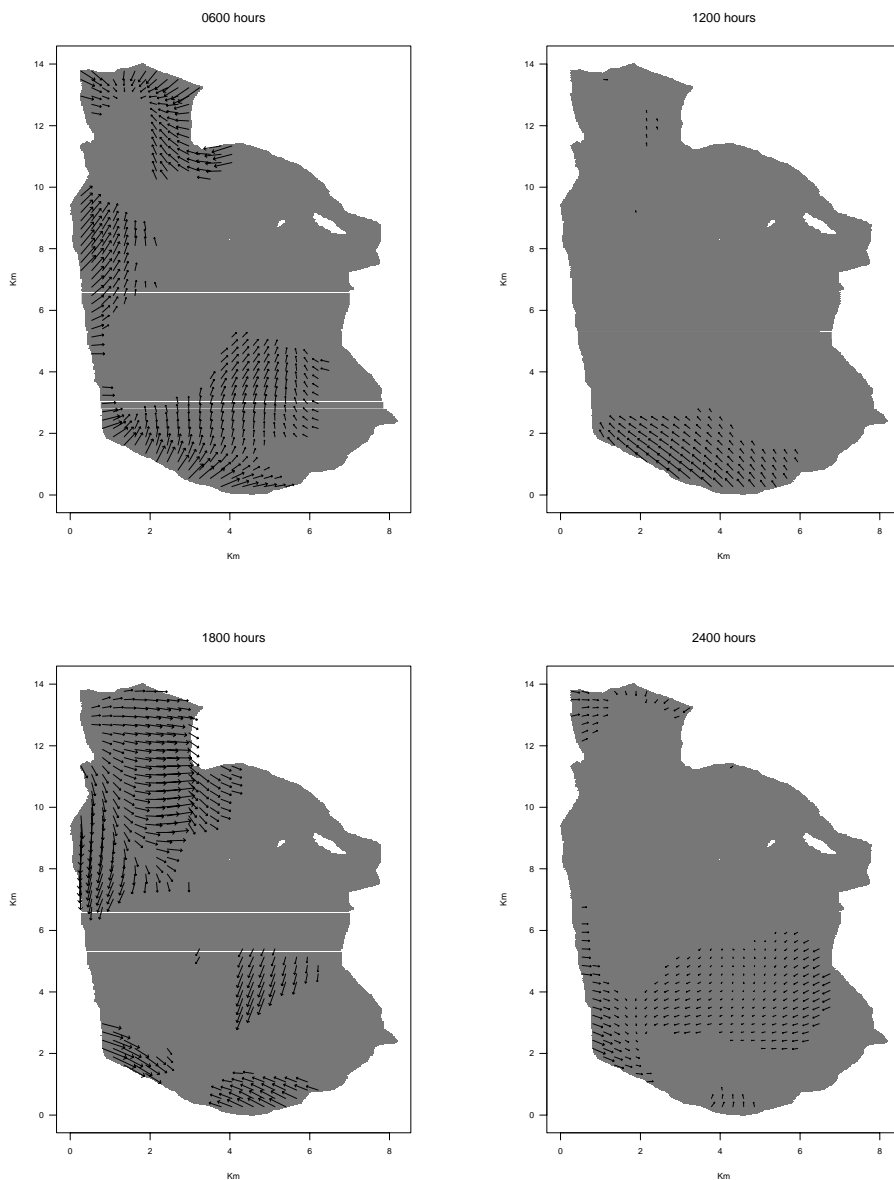
Figure 7: Locations where absolute value of the t-statistic exceeds the 95% level, based on jackknife computations.

locations. The elk appear to be more active at 0600 and 1800 hours, but staying in a local area at 1200 and 2400 hours. This is consistent with the information of Figure 4. Further one sees regions of the vector fields converging towards areas of attraction, which may be compared with the hotspots of Figure 5.

Further discussion of sampling uncertainty is needed in order to formalize the inferences. Concerning the vector field plots the jackknife, Mosteller and Tukey [20], was employed. In its implementation 50 of the 53 elk trajectories were used, 5 trajectories were dropped each time in the evaluations of 10 pseudo estimates. Figure 7 graphs the locations where the absolute values of t-statistics exceed the 95 percent point of the Student-t distribution with 9 degrees of freedom. The elk remain most mobile around 0600 and 1800 and there remain the suggestions of points of convergence.

These plots provide insights into diel patterns of elk movements. The results presented in these four panels are consistent with the previous result that the time of day effect is not simply additive. Figure 6 remains important because the lengths of the arrows provide estimates of the animals' speeds as a function of location.

In summary, this section has presented a method for estimating the vector field $\mu(x, y, t)$ when it is smooth in location and time of day. The dependence on time of day and location does not appear additive.

6.3   Modelling Σ

Residuals are an important tool for seeking omitted variables, for inferring non-linearities in entered variables and for learning about the basic variability in the model. In the model (2.3) variability is represented by the term $\Sigma(\mathbf{r}, t)d\mathbf{B}(t)$. In this section residuals of the general fit to time of day and location are employed.

The estimated variance-covariance matrix of the x- and y-residuals is

$$\begin{bmatrix} .05272 & .00028 \\ .00028 & .06301 \end{bmatrix}$$

It is near diagonal, consistent with the assumption of the independence of the x- and y-components of the Brownian noise of (2.3).

When an analysis of variance is carried out on the log residuals-squared the P-values are all negligeable consistent with the variance needing to be modelled as depending nonlinearly on both time of day and location.

6.4   Assessment of fit

It is necessary to assess the goodness of fit of the model fit before drawing substantial conclusions. For example the P-values in the ANOVA tables are based on gaussian, white-noise assumptions. One difficulty in studying the goodness of fit is that the data involved are unequally-spaced in time. The periodogram is a useful statistic to employ in such a situation, see Brillinger [5].

Let

$$\hat{\epsilon}_m(t_{m,k}) = \frac{(X(t_{m,k+1}) - X(t_{m,k}) - \hat{\mu}(\mathbf{r}_{m,k}, \langle t_{m,k} \rangle))(t_{m,k+1} - t_{m,k}))}{\sqrt{\hat{\sigma}(\mathbf{r}_{m,k}, \langle t_{m,k} \rangle)}}$$

denote the standardized x-residuals in the case of the m-th elk Consider the empirical Fourier transform

$$d_m{}^T(\lambda) = \sum_k \hat{\epsilon}_m(t_{m,k}) \exp\{-i\lambda t_{m,k}\}$$

evalutated by summing over the available time points, $t_{m,k}$, for the m-th animal. The x-periodogram is defined by

$$|d_m{}^T(\lambda)|^2$$

This statistic was computed for each of the $M = 53$ elk and the results averaged. Assuming the series have common power spectrum its distribution is approximately a multiple of chi-squared on $2*M$ degrees of freedom see Brillinger [6]. In the case that the spectrum is constant, the periodograms values will fluctuate approximately about a constant level. Figure 8 provides the average of the 53 periodogams and approximate 95% marginal confidence limits about the mean periodogram value. The top panel refers to the x-values while the middle panel refers to the y-values. The high values at the low frequencies provide some evidence that the residuals are not white noise. A time series model allowing weak correlation between values nearby in time might be necessary. The high values at a frequency of 1 cycle/day suggest that the time of day component has not been completely removed. This may be because its shape is changing during the season.

The bottom panel provides the estimated coherence between the x- and y-components and an approximate upper 95% null point. It is consistent with the components being independent as in (4.1).

Figure 9 addresses the issue of the independence of the trajectories of the elk when time of day and location effects have been "removed" by computing
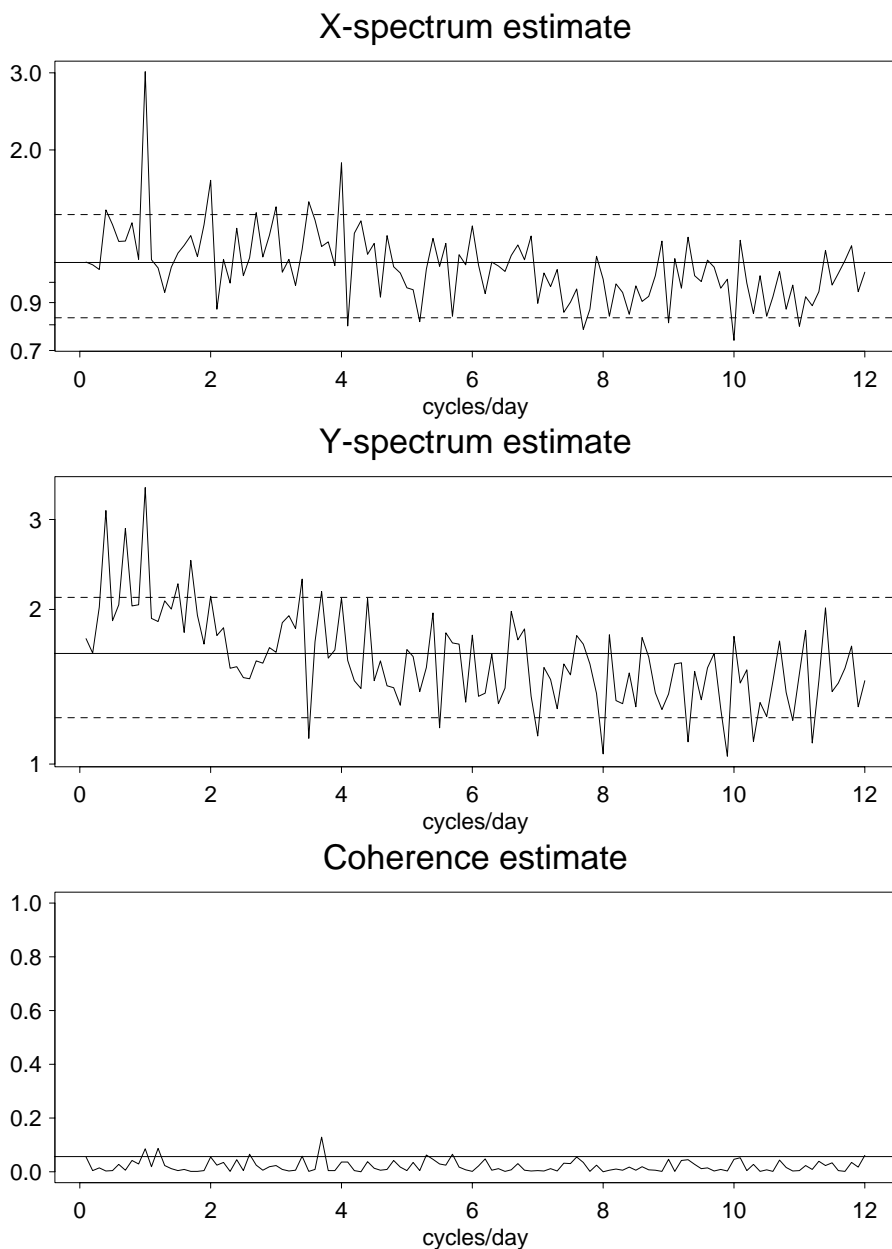
Figure 8: The $x-$ and $y-$periodograms of the standardized residuals and the estimated coherence. The horizontal line in the coherence plot is the upper 95% null point.

F ratio for X-values

cycles/day

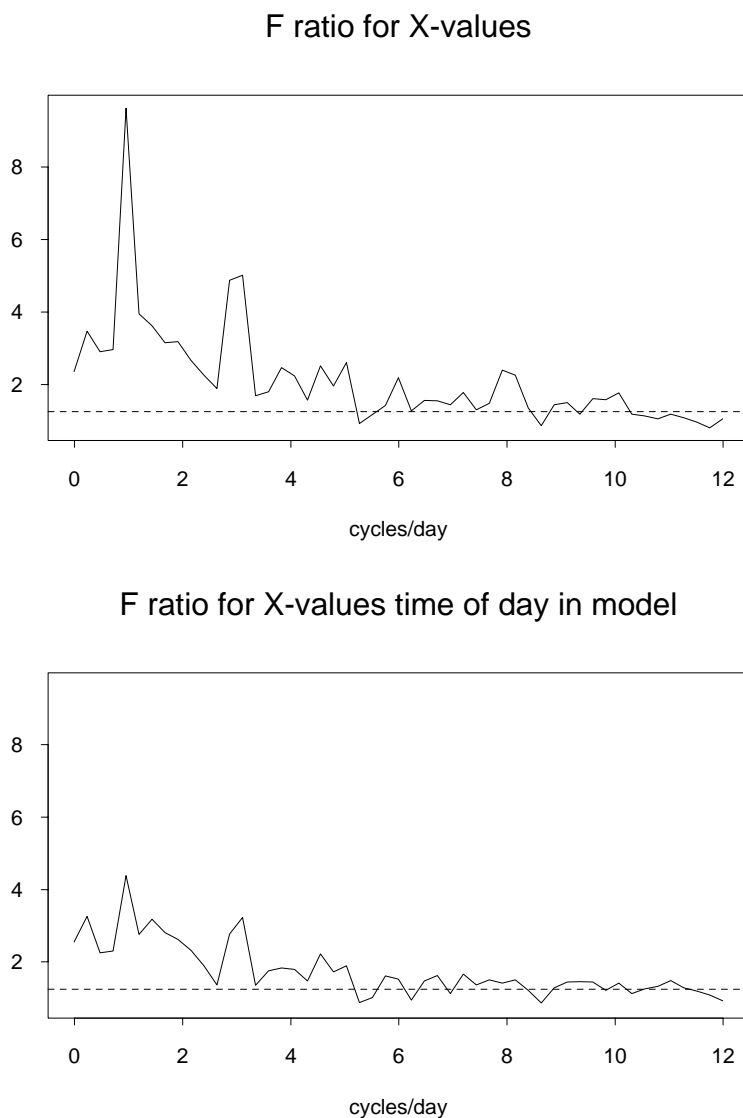F ratio for X-values time of day in model

cycles/day

Figure 9: The X-component F-ratios for looking for a common effect amongst the elk. The horizontal lines in the plots are the upper 95% null point of the F-statistic.

the standardized residuals. An elementary model introducing equicoherence amongst the elk residuals contains a component common to all the elk and independent superposed noise. This leads to the following analysis of variance approach. Consider frequencies λ_l near a frequency λ of interest. For the Fourier transform of the m-th elk write

$$d_m{}^T(\lambda_l) = \mu + \alpha_l + \eta_{lm}$$

with μ a mean level, the α_l random effects, and the η's errors. For a given λ this is a classic one-way layout and the presence or absence of the α's may be studied by an F-statistic. The results are given in Figure 9, which are plots of the F-statistics as functions of frequency. Results are presented for the x-component. The top panel graphs the F-statistic having removed only the effects of location, i.e., time of day effect is not removed. One sees peaks at the frequency of one cycle/day and its second harmonic. The lower panel shows the statistic having fit both location and time of day. In each case the horizontal dashed line is the upper 95% level of the null F distribution. The empirical values are seen to be fluctuating about this level, however the peaks are much reduced. There is evidence for some dependence at the low frequencies. Despite that, the results of the jackknife computations are still pertinent. Modelling the remaining dependence is a problem for future attention.

One might examine the possibility that the drift function corresponds to a potential function. This was done in Brillinger et al. [6], separately for the day and night data. The hypothesis could not be rejected.

7   Discussion and summary

This paper has investigated the use of stochastic differential equations to model wildlife motion. Models of movement are useful tools to study the ecology of animal behavior and test ideas concerning foraging strategies, habitat preferences, and the dynamics of population densities. Specific questions for large animals like elk include: What are the effects of phenomena such as roads, cover, forage, time of day, season, and human disturbance? How should one allocate forage amongst wild and domestic species? What is the effect of vehicular traffic? Is change taking place? What is the sequence of habitat use? Understanding the physical and biological mechanisms that regulate animal movements is clearly a complex problem. Formal models seem both useful and necessary.

The analytic techniques of potential function, stochastic differential equations, and nonparametric estimation were employed in this paper in the model development. The assumption of a potential function led to the setting down of a

stochastic differential equation for a diffusion process. This SDE assumption further motivated the estimates computed. It may be remarked that diffusion processes are Markov, whereas more realistic equations would involve time lags and the process therefore not be Markov. The statistics presented in Figures 8 and 9 provide possible evidence of this occurrence.

8   Some probability problems

1. That the ringed-seal is constrained in a lake and that there is a fence around the elk reserve is basic. One would like to fit probability models that recognize this.

Question 1. Given the diffusion process (2.3), how does one tell from the form of μ and Σ that there is a closed boundary that keeps the process inside once it starts there?

2. For some purposes, e.g. understanding variability, one would like to simulate realizations of the process. The boundary behavior would need to be defined.

Question 2. How does one simulate trajectories of a diffusion process restricted to stay in a region?

3. Sometimes an elk is seen to wander along the fence, i.e. not be reflected back.

Question 3. How does one include in the model the possibility that the process may follow the boundary for a period?

4. The previous questions had in mind a process in space. Consider the case of a process on the line.

Question 4. For a real-valued process how does one tell that its sample paths are non-negative?

References

[1]   Banon, G. and Nguyen, H. T. Recursive estimation in diffusion model. *SIAM J. Control and Optimization* **19** (1981), 676–685.

[2] Bertrand, P. Comparison de l'erreur quadratique moyenne intégrée pour différents estimateurs du coefficient de diffusion d'un processus. *C. R. Acad. Sci. Paris* **327** (1998), 399–404.

[3] Bhattacharya, R. N. and Waymire, E. *Stochastic Processes with Applications.* New York, Wiley (1990).

[4] Brillinger, D. R. A particle migrating randomly on a sphere. *J. Theoretical Probability* **10** (1997), 429–443.

[5] Brillinger, D. R. Examining an irregularly sample time series for whiteness, *Resenhas*. **4** (2000), 423–431.

[6] Brillinger, D. R., Preisler, H. K., Ager, A. A. and Kie, J. G. The use of potential functions in modelling animal movement. Pp. 369–386 in *Data Analysis from Statistical Foundations.* (Ed. A. K. Md. E. Saleh.) Nova Science, New York (2001).

[7] Brillinger, D. R. and Stewart, B. S. Elephant seal movements: modelling migration. *Canadian J. Statistics* **26** (1998), 431–443.

[8] Burgière, P. Théorème de limite centrale pour un estimateur non paramétrique de la variance d'un processus diffusion multidimensionelle. *Annales de l'Institut Henri Poincaré, Section B, Calcul des Probabilities et Statistique* **29** (1993), 357–389.

[9] Cleveland, W. S., Grosse, E. and Shyu, W. M. Local regression models. Pp. 309–376 in *Statistical Models in S* (Eds. J. M. Chambers and T. J. Hastie). Pacific Grove, Wadsworth (1992).

[10] Dohnal, G. On estimating the diffusion coefficient. *Journal of Applied Probability* **24** (1987), 105–114.

[11] Dunn, J. E. and Brisbin, I. L. Characterization of the multivariate Ornstein-Uhlenbeck diffusion process in the context of home range analysis. 181–205 in *Statistical Theory and Data Analysis.* (Ed. K. Matusita). B.V. North-Holland, Elsevier (1985).

[12] Dunn, J. E. and Gipson, P. S. Analysis of radio telemetry data in studies of home range. *Biometrics* **33** (1977), 85–101.

[13] Genot-Catalot, V., Laredo, C., Picard, D. Nonparametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics* **19** (1992), 317–335.

[14] Goldstein, H. *Classical Mechanics*. Addison-Wesley, New York (1950).

[15] Hastie, T. J. Generalized additive models. Pp. 195–247 in *Statistical Models in S* (Eds. J. M. Chambers and T. J. Hastie). Pacific Grove, Wadsworth (1992).

[16] Hastie, T. J. and Tibshirani, R. J. *Generalized Linear Models.* London, Chapman and Hall (1990).

[17] Karlin, S. and Taylor, H.M. *A Second Course in Stochastic Processes*. New York, Academic (1981).

[18] Kie, J. G. and R. T. Bowyer. Sexual segregation in white-tailed deer: density dependent changes in use of space, habitat selection, and dietary niche. *Journal of Mammalogy* **80** (1999), 1004–1020.

[19] Kloeden, P. E. and Platen, E. *Numerical Solution of Stochastic Differential Equations.* New York, Springer (1995).

[20] Mosteller, F. and Tukey, J. W. *Regression a Second Course*. Addison-Wesley, Reading (1977).

[21] Nelson, E. *Dynamical Theories of Brownian Motion*. Princeton: Princeton U. Press (1967).

[22] Prakasa Rao, B. L. S. *Statistical Inference for Diffusion Type Processes.* Arnold, London (1999).

[23] Preisler, H. K. and Akers, P. Autoregressive-type models for the analysis of bark beetle tracks. *Biometrics* **51** (1995), 259–267.

[24] Preisler, H. K., Brillinger, D. R., Ager, A. A. and Kie, J. G. Analysis of animal movement using telemetry and GIS data. *Proc. ASA Section on Statistics and the Environment*, (1999), 100–105.

[25] Preisler, H. K., Brillinger, D. R., Ager, A. A., Kie, J. G. and Akers, R. P. (2001). Stochastic differential equations: a tool for studying animal movement. *Proc. IUFRO 4.11 Conference, Greenwich 26-29 June, 2001*.

[26] Prohorov, Yu. V. and Rozanov, Yu. A. *Probability Theory*. Springer-Verlag, New York (1969).

[27] Rowland, M. M., Bryant, L. D., Johnson, B. K., Noyes, J. H., Wisdom, M. J. and Thomas, J. W. *The Starkey Project: History, Facilities, and Data Collection Methods for Ungulate Research*. Technical Report PNW-GTR-396, Forest Service, USDA (1997).

[28] Rowland, M. M., M. J. Wisdom B. K. Johnson and J. G. Kie. Elk distribution in relation to roads. *The Journal of Wildlife Management* **64** (2000), 672–685.

[29] Sorensen, M. Estimating functions for discretely observed diffusions: a review. Pp. 305–326 in *Selected Proceedings of the Symposium on Estimating Functions*. Eds. Basawa, I. V, Godambe, V. P. and Taylor, R. L. Vol. 32 Lecture Notes. Hayward, Institute of Mathematical Statistics (1997).

[30] Stewart, B. S. Ontogeny of differential migration and sexual segregation in northern elephant seals. *Journal of Mammalogy* **78** (1997), 1101–1116.

[31] Stewart, J. *Calculus, Early Transcendentals*. Brooks/Cole, Pacific Grove (1991).

[32] Turchin, P. *Quantitative Analysis of Movement*. Sunderland, Sinauer (1998).

[33] White, G. C. and Garrott, R. A. *Analysis of Wildlife Radio-tracking Data.* Academic Press, New York (1990).

David R. Brillinger
Statistics Department, University of California
Berkeley, CA 94720, U.S.A.
E-mail: brill@stat.berkeley.edu

Haiganoush K. Preisler
Pacific Southwest Research Station, USDA Forest Service
Albany, CA 94710, U.S.A.
E-mail: hpreisler@fs.fed.us

Alan A. Ager, John G. Kie
Pacific Northwest Research Station, USDA Forest Service
La Grande, OR 97850, U.S.A.
E-mail: aager@fs.fed.us, jkie@fs.fed.us

Brent S. Stewart
Hubbs-Sea World Research Insitute
San Diego, CA 92109, U.S.A.
E-mail: stewartb@hswri.org