



Available at

www.ElsevierMathematics.com

POWERED BY SCIENCE @ DIRECT®

Journal of Statistical Planning and
Inference 122 (2004) 43–63

journal of
statistical planning
and inference

www.elsevier.com/locate/jspi

An exploratory data analysis (EDA) of the paths of moving animals

David R. Brillinger^{a,*}, Haiganoush K. Preisler^b,
Alan A. Ager^c, John G. Kie^c

^a*Statistics Department, University of California, 367 Evans Hall 3860, Berkeley, CA 94720, USA*

^b*Pacific Southwest Research Station, USDA Forest Service, Albany, CA 94710, USA*

^c*Pacific Northwest Research Station, USDA Forest Service, La Grande, OR 97850, USA*

Accepted 15 June 2003

Abstract

This work presents an exploratory data analysis of the trajectories of deer and elk moving about in the Starkey Experimental Forest and Range in eastern Oregon. The animals' movements may be affected by habitat variables and the behavior of the other animals. In the work of this paper a stochastic differential equation-based model is developed in successive stages. Equations of motion are set down motivated by corresponding equations of physics. Functional parameters appearing in the equations are estimated nonparametrically and plots of vector fields of animal movements are prepared. Residuals are used to look for interactions amongst the movements of the animals. There are exploratory analyses of various sorts. Statistical inferences are based on Fourier transforms of the data, which are unequally spaced. The sections of the paper start with motivating quotes and aphorisms from the writings of John W. Tukey.

© 2003 Elsevier B.V. All rights reserved.

MSC: primary 62M10, 62M15, 60H10; secondary 37N25

Keywords: Dependence; Deer; Elk; Exploratory data analysis; Fourier analysis; Paths; Starkey Experimental Forest; Time series; John W. Tukey; Vector field plots

1. Introduction

The first task of the analyst of data is quantitative detective work, ... His second task is to extract as clearly as possible what the data says about certain specified

* Corresponding author. Tel.: +1-510-642-2781; fax: +1-510-642-7892.

E-mail addresses: brill@stat.berkeley.edu (D.R. Brillinger), hpreisler@fs.fed.us (H.K. Preisler), aager@fs.fed.us (A.A. Ager), jkie@fs.fed.us (J.G. Kie).

parameters. His later tasks are to assess the contributions of these statements from all causes, systematic or whatever ... Often the purpose of good analysis is not so much to do well in catching what you want but rather to do well ... in rejecting what you don't want. (Tukey, 1971)

The problem of interest this work presents is the description of the movement of free-ranging animals. In particular, the paths of radio-collared Rocky Mountain Elk (*Cervus elaphus*) and mule deer (*Odocoileus hemionus*) moving in the 7700 ha. Starkey Experimental Forest are studied. The details of the experiment are provided in Rowland et al. (1997). Models of movement are useful tools to study the ecology of animal behavior and to test ideas concerning foraging strategies, habitat preferences, and the dynamics of population densities. The description of movement is a complex problem owing to the array of biological and physical mechanisms that control how animals move on large landscapes.

Questions of interest include: How to describe the deer movements? How to describe the elk movements? How do their movements differ? How do the animals interact? Are there some surprises?

Various exploratory data analyses have been carried out in the course of the work, but only the particularly enlightening ones have been presented. A model of a stochastic differential equation (SDE) driven by a process with stationary increments is developed. At the outset Brownian motion was considered for the driving process. However, this would have implied that the solution of the SDE was Markov and the analysis as well as the biology contradicted this possibility.

2. A model

Data analysts regard their models as a basis from which to measure deviation, as a convenient bench mark in the wilderness, expecting little truth and relying on less. (Tukey, 1979)

The approach developed in the work is to assume that the animals are moving in accordance with bivariate stochastic differential equations of the form

$$d\mathbf{r}(t) = \mu\{\mathbf{r}(t), t\} dt + \Sigma\{\mathbf{r}(t), t\} d\mathbf{V}(t) \quad (1)$$

with $\mathbf{r}(t)$ a 2-vector representing the location of an animal of concern at time t . Here μ and Σ are parameters and \mathbf{V} is a random function driving the movement. The parameters and the process \mathbf{V} control the direction and speed of motion. With $E\{d\mathbf{V}(t) | \mathbf{r}(t)\} = \mathbf{0}$ and $\text{var}\{d\mathbf{V}(t) | \mathbf{r}(t)\} = \mathbf{I} dt$ one has the interpretations

$$E\{d\mathbf{r}(t) | \mathbf{r}(t)\} = \mu\{\mathbf{r}(t), t\} dt \quad \text{and} \quad \text{var}\{d\mathbf{r}(t) | \mathbf{r}(t)\} = \Sigma\{\mathbf{r}(t), t\} \Sigma\{\mathbf{r}(t), t\}' dt.$$

Later in the paper it will be assumed that the process \mathbf{V} has stationary increments. Commonly it is assumed that \mathbf{V} is a Brownian or a Lévy process, but this would imply that the process $\mathbf{r}(t)$ is Markov.

There are explanatory variables to include in the model. The permanent spatial ones, such as the fence and foraging areas around the reserve, lie in the functions $\mu(\cdot)$ and $\Sigma(\cdot)$.

3. The data

You must ‘sit loose’ to data, to results of analyzing data, and to interpretations of these results, if you are to get full value of any of them. (Tukey, 1961)

Elk and telemetry data of the period 9 April to 15 May of the years 1993, 1995, 1996 are studied. The data from 1994 were studied in Preisler et al. (1999, 2001), Brillinger et al. (2001a,b, 2002). The numbers of deer in 3 years are 5, 30, 8, respectively, while the corresponding elk numbers are 30, 19, 45.

The data are spatial-temporal. The animals are labeled by $m = 1, \dots, M$, and their locations are recorded at times, t_{mk} , $k = 1, \dots, K_m$ for the m th animal. Also available are various explanatory variables describing forest vegetation and topography suspected to influence animal movement. Other habitat features such as distance to road and distance to hiding cover may be derived. The locations are denoted $\mathbf{r}_m(t) = \{X_m(t_{mk}), Y_m(t_{mk})\}$, corresponding to the Universal Transverse Mercator (UTM) coordinates of the k th time measurement of the m th animal.

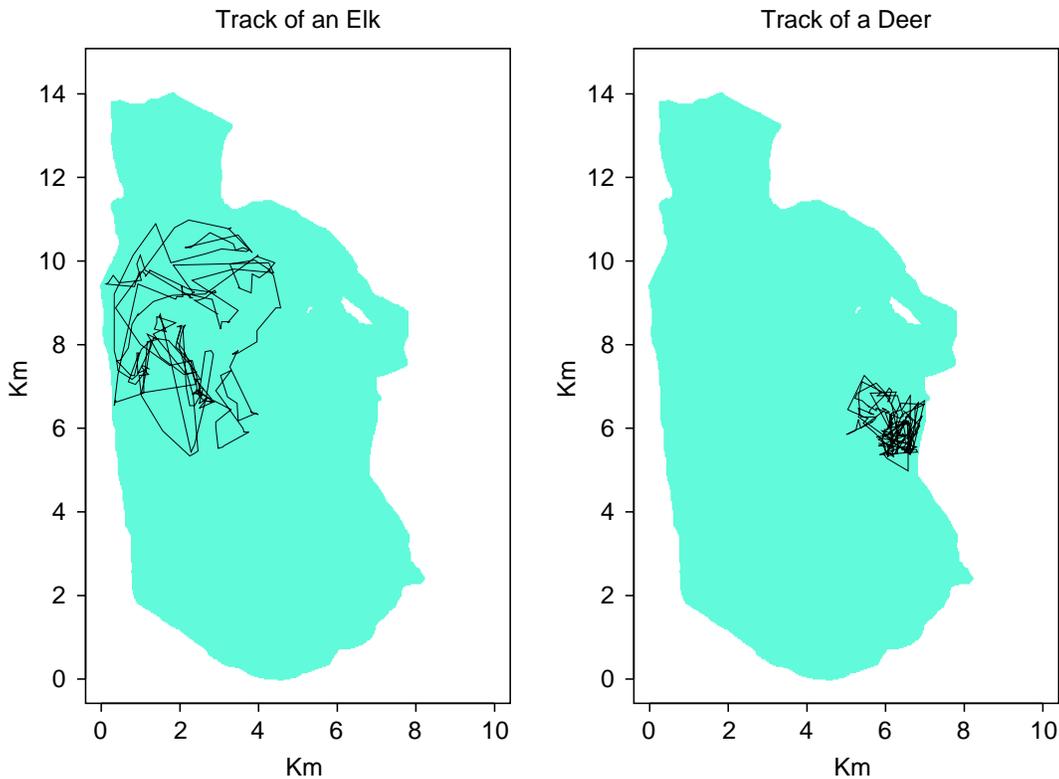


Fig. 1. Elk and deer tracks for 1 month. The paths are sampled in time, hence the straight line segments.

Fig. 1 provides examples of the tracks of one of the deer and of one of the elk. One sees that these two particular animals did not move too much over Starkey, but in studies of the elk over a whole summer period, Preisler et al. (1999, 2001), Brillinger et al. (2001a,b, 2002), some of the elk are found to visit widely spread locations.

4. EDA

... I hope that I have shown that exploratory data analysis is actively incisive rather than passively descriptive, with a real emphasis on the discovery of the unexpected ... (Tukey, 1973)

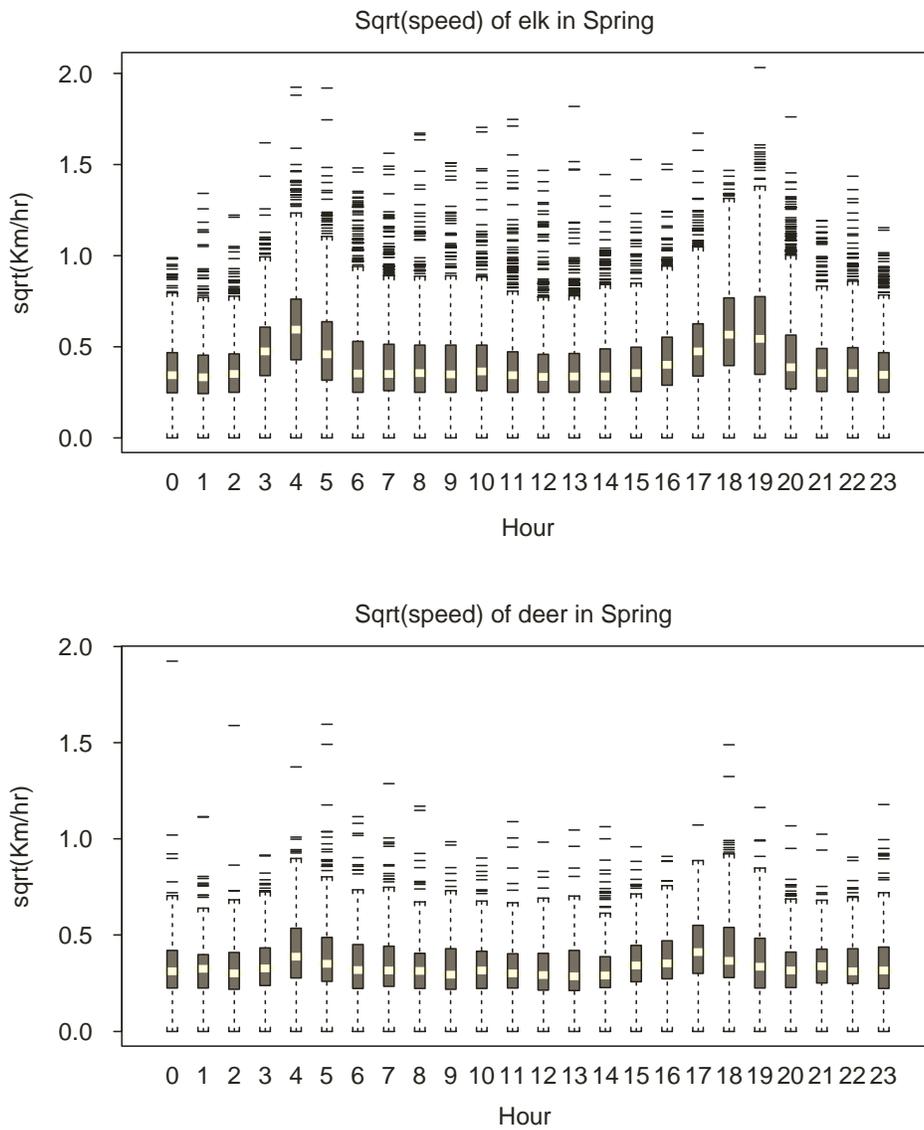


Fig. 2. Parallel boxplots of the square roots of the elk and deer speeds by hour of the day.

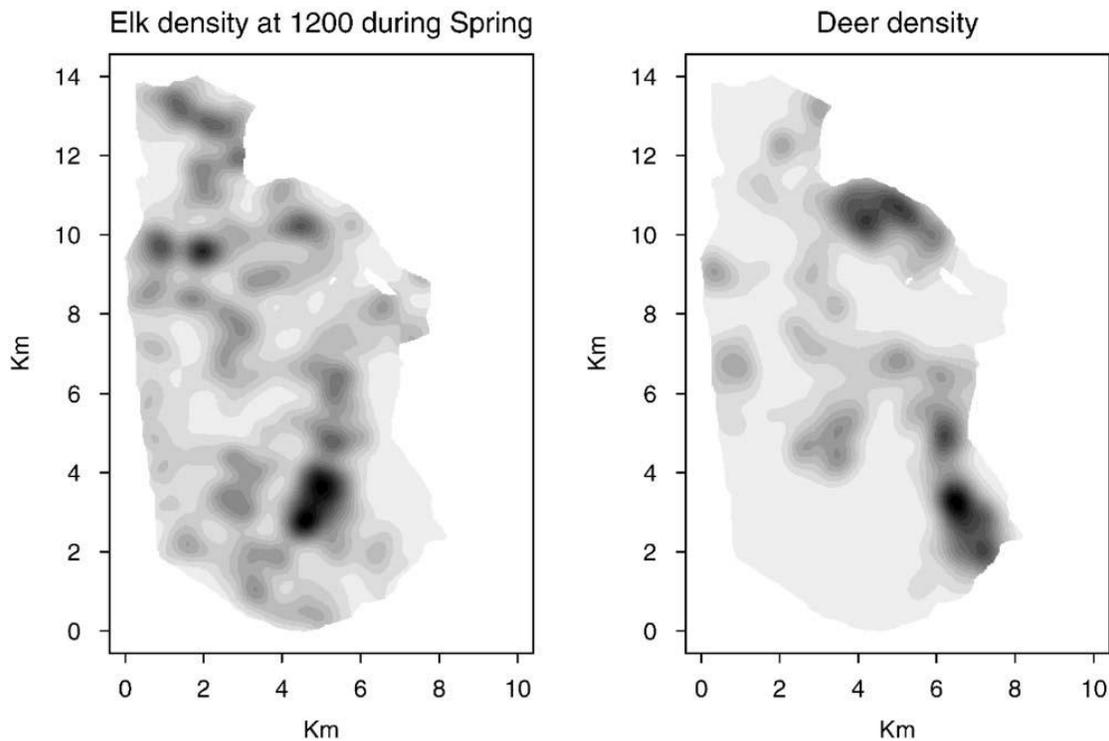


Fig. 3. Estimated location densities for the elk and the deer at noon.

Many good techniques are simple and direct, with transparent relevance, for example the box-plot and the scatter-plot. (Mallows and Tukey, 1982)

To begin the analysis some time domain plots are provided and then some spatial ones. Fig. 2 is a parallel boxplot of the square roots of estimated elk and deer speeds by hour of the day. The groups of animals appear substantially more mobile around 0500 and 1800 h and less active at night and mid-day, with the deer somewhat less mobile. These observations are consistent with previous studies that have described strong dawn and dusk movements by elk and deer between foraging areas and hiding cover.

In preparing these plots animal speed was estimated by dividing the distance between two successive locations by the difference of the corresponding observation times. For stability the observations with time differences between 0.08 and 2.5 h were used. It was felt that this gave some stability to the results without biasing them overwhelmingly. With further assumptions the bias might be corrected for to a degree. The square root re-expression is employed in the plots to de-emphasize the effects of stragglers.

Fig. 3 provides kernel density estimates of the animals' noon locations based on all the data available. Noon was picked as, following Fig. 2, the animals' were less mobile then. There are different hot spots, i.e. locations of congregation, for the deer and the elk. There is a large area in the south-west that the deer appear to avoid.

5. Fourier analysis

Though I was unaware (or slightly aware) of it at the time, it is now clear to me that spectrum analysis, with its challenging combination of amplified difficulties and forcible attention to reality, has done more than any other area to develop my overall views of data analysis. (Tukey, 1984, p. xxxix)

According to model (1)

$$\dot{\mathbf{r}} = \mu\{\mathbf{r}(t), t\} + \Sigma\{\mathbf{r}(t), t\}\dot{\mathbf{V}}$$

and an animal's velocity depends on location and time. Suppose, to motivate the analyses of this section, the velocity process, $\dot{\mathbf{r}}$ is stationary, as would be the case if the μ and Σ did not depend on t and $\dot{\mathbf{V}}$ was stationary.

Fig. 4 provides plots of the X and Y coordinates of the velocities $(\mathbf{r}(t_{m,k+1}) - \mathbf{r}(t_{mk})) / (t_{m,k+1} - t_{mk})$ at the time points t_{mk} , for the elk of Fig. 1. There is a gap when the measurements were not collected. Despite some apparent reduced variability at the outset acting as if the processes are stationary does not appear unreasonable.

In the analyses to come the following Fourier transform of the process is employed

$$d^T(\lambda) = \int_0^T h(t)\dot{\mathbf{r}}(t) \exp\{-i\lambda t\} dt. \quad (2)$$

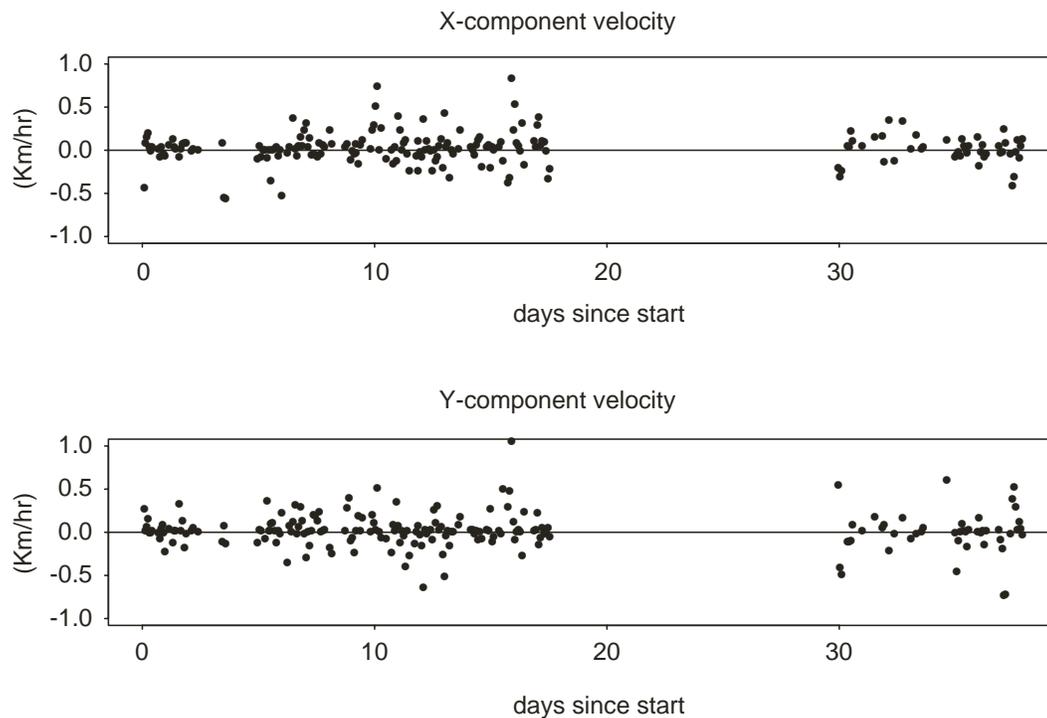


Fig. 4. Estimated velocities in the X and Y directions for one of the elks as a function of time.

Here $-\infty < \lambda < \infty$ and $h(t)$ a taper function to handle gaps in the time period of data collection and to control leakage as necessary. This Fourier transform has convenient large sample statistical properties in the case of a stationary mixing process, see Appendix B. For the m th animal it may be approximated by

$$\mathbf{d}_m^T(\lambda) = \sum_k h(t_{mk})(\mathbf{r}_m(t_{m,k+1}) - \mathbf{r}_m(t_{mk})) \exp\{-i\lambda t_{mk}\} \tag{3}$$

for animal m . In many stationary cases such Fourier transforms are approximately Gaussian and independent for distinct frequencies λ . The periodogram matrix at frequency λ is now defined as

$$\frac{1}{2\pi \int h(t)^2 dt} \mathbf{d}_m^T(\lambda) \overline{\mathbf{d}_m^T(\lambda)'}.$$

This periodogram was computed for each animal and year. Assuming the processes have a common spectral density matrix, one can use averages over the individual periodograms as an estimate of it. The estimated power spectra are then, approximately, multiples of chi-squared. Were the series \mathbf{r}_m white noise the spectrum would be constant, and the periodograms' values would fluctuate about a constant level see Appendix B.

Figs. 5 and 6 provide the results for the elk and deer, respectively. Also, approximate 95% marginal confidence limits about the mean periodogram values are provided by the dashed lines. The bottom panel provides the estimated coherence and an approximate upper 95% null level.

For a bivariate stationary time series (X, Y) with a sampling spacing of 1 the coherency may be defined as

$$\lim_{T \rightarrow \infty} E\{d_X^T(\lambda) \overline{d_Y^T(\lambda)}\} / \sqrt{E\{|d_X^T(\lambda)|^2\} E\{|d_Y^T(\lambda)|^2\}}.$$

The coherence is the modulus-squared of the coherency.

In the case of the elk one notes the presence of strong peaks corresponding to a circadian rhythm of period 1 day. (The vertical axis is on a log scale.) Such a rhythm was suggested by Fig. 2. In the case of the Y -component, for both the elk and the deer, there is a fall-off in power with increasing frequency. This suggests autocorrelation beyond that of the 24 h period and complicates the interpretation of the estimate. The result for the X -component is consistent with white noise. (One reason for this is that because of the physical locations of the LORAN transmitters and receivers used to estimate the locations there is much more noise in the estimated value of the X -coordinate.)

Looking at the bottom panels of Figs. 5 and 6, the coherence estimate suggests that the X and Y components are not related in a linear time invariant manner. This would follow for the process (1) were the Σ appearing diagonal and the components of the process \mathbf{V} unrelated.

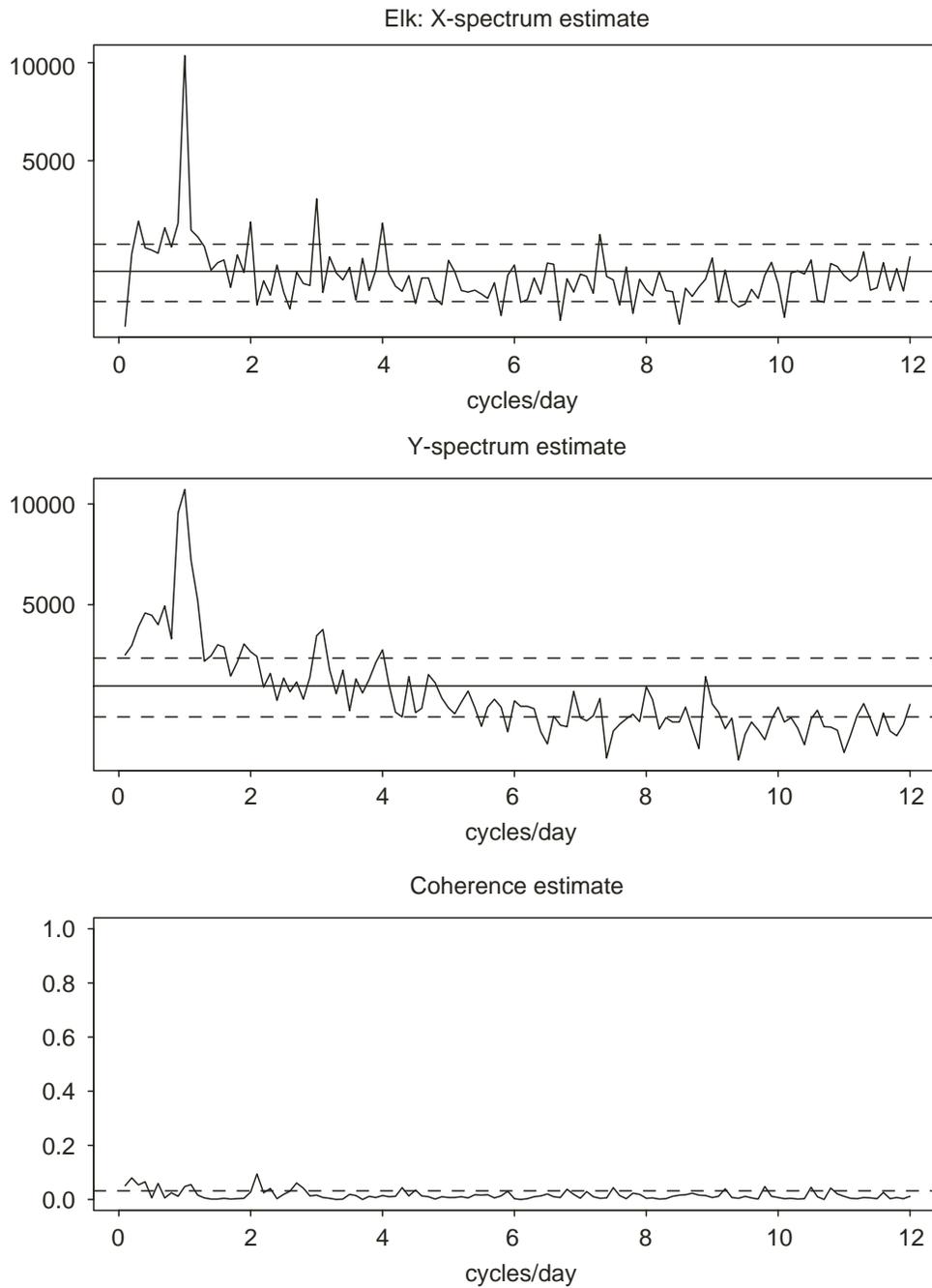


Fig. 5. Estimates of the spectrum of the X and Y coordinates and their coherence for the elk. The dashed lines provide approximate 95% bounds.

The models to be employed subsequently will include a 24 h period, specifically t of (1) will be replaced by $\langle t \rangle$, the time of day, in μ and Σ .

The computations leading to Figs. 5 and 6 were carried out using expression (3) directly. It is worth noting that JWT had some ideas on how to compute an

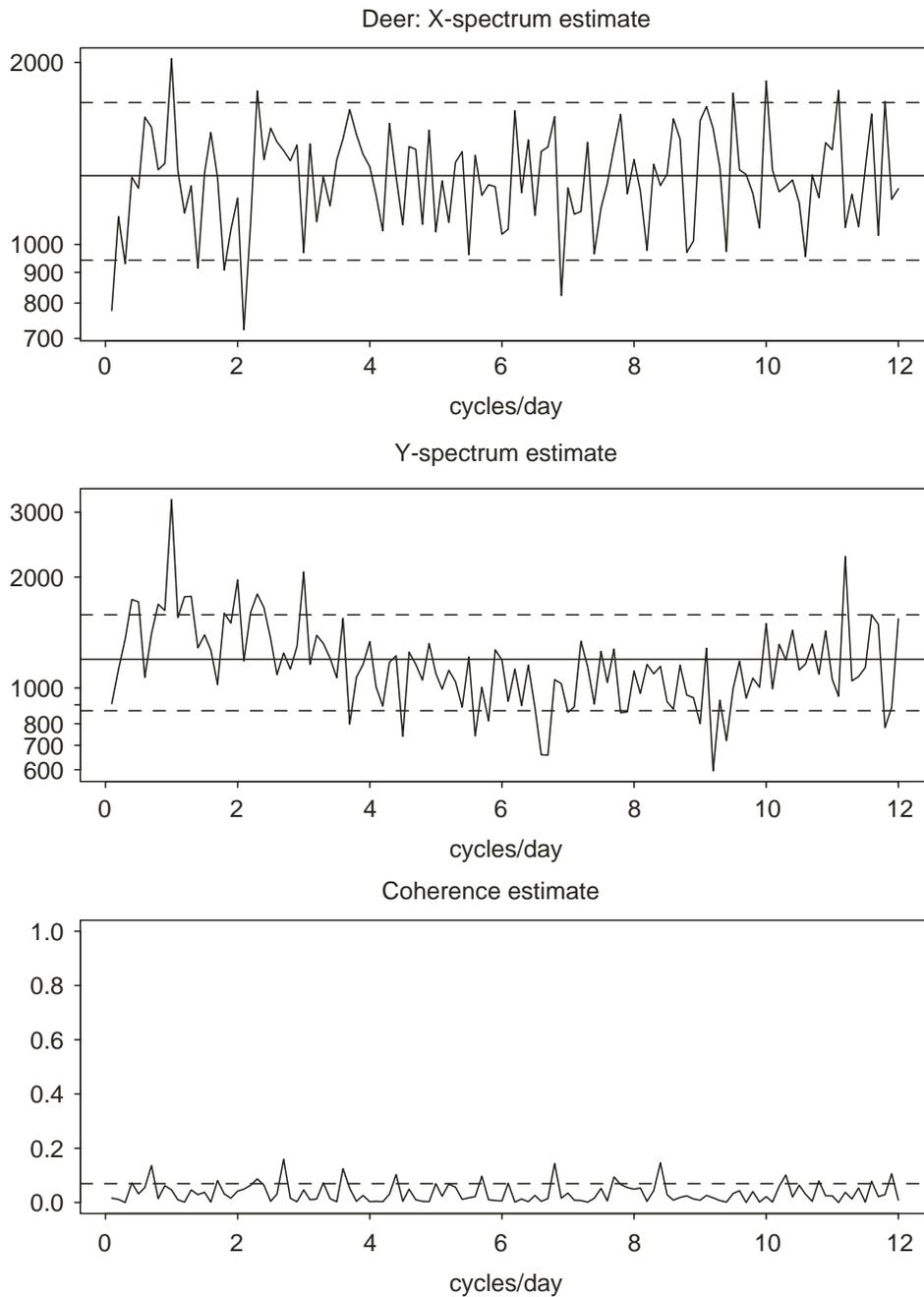


Fig. 6. Spectrum estimates for the deer.

approximate FFT for irregularly spaced data, see his letter reproduced in Appendix A. (A recent reference is [Dutt and Rokhlin \(1993\)](#).) The basis for the approximate 95% bounds in the figures is the approximate normality of the Fourier transform (3), see Appendix B.

6. Smoothing

We need techniques that will help us see what the data are saying, even when they don't follow a straight line. (Tukey, 1977, p. 207)

The main tasks of pictures are then: to reveal the unexpected, to make the complex easier to perceive. Either may be effective for that which is important above all: suggesting the next step in analysis, or offering the next insight. (Tukey, 1975)

There are a variety of methods to approximate SDEs, Kloeden and Platen (1995). A naive approximation to (1) is provided by writing

$$\{\mathbf{r}(t_{k+1}) - \mathbf{r}(t_k)\} / (t_{k+1} - t_k) \approx \boldsymbol{\mu}\{\mathbf{r}(t_k), t_k\} + \boldsymbol{\Sigma}\{\mathbf{r}(t_k), t_k\} \mathbf{Z}_k / \sqrt{t_{k+1} - t_k} \quad (4)$$

$k=1, 2, \dots$ with $t_1 < t_2 < t_3 < \dots$ observation times and the \mathbf{Z}_k satisfying $E\{\mathbf{Z}_k | \mathbf{r}(t_k)\} = \mathbf{0}$, $\text{var}\{\mathbf{Z}_k | \mathbf{r}(t_k)\} = \mathbf{I}$.

In terms of the individual components (X, Y) of \mathbf{r} one has been led to the model

$$\frac{\Delta X}{\Delta t} = \mu_x(X, Y, \langle t \rangle) + \text{noise}, \quad \frac{\Delta Y}{\Delta t} = \mu_y(X, Y, \langle t \rangle) + \text{noise}. \quad (5)$$

If the drift function components, $\mu_x(\cdot)$, $\mu_y(\cdot)$, are smooth and unknown, one has a nonparametric regression problem. Here estimates of the functions were calculated via the function `gam()` of Hastie (1992) making use of the function `lo()` of Cleveland et al. (1992). Following (4) the weights $t_{m,k+1} - t_{m,k}$ were used for the m th animal and its k th difference.

Analyses are done separately for the deer and elk and are presented in Fig. 7 for the time 0500. This time was selected because, following Fig. 2, the animals were particularly active then. As indicated previously the observations with $0.08 < t_{k+1} - t_k < 2.5$ h are used in order that the instantaneous velocities may be approximated reasonably. The plots are in the form of vector-fields with the length of an arrow is proportional to the estimated speed at that location. Its angle corresponds to the estimated direction of movement. The directions of arrows at different times of day may be compared with maps of habitat showing features such as locations of roads, streams or foraging areas. The figure may also be compared with Fig. 3. The movement vectors at 0500 h (Fig. 7) may also be used with the density estimate at 1200 h (Fig. 3) to visualize the morning movements of the elk and deer. All three variables, x , y , $\langle t \rangle$, in the drift term appear significant implying the dependence of drift on location and time of day.

It may be useful to plot only arrows whose lengths are "significant". This is done in Brillinger et al. (2001a) with uncertainties estimated by a jackknife procedure that involves dropping groups of animals.

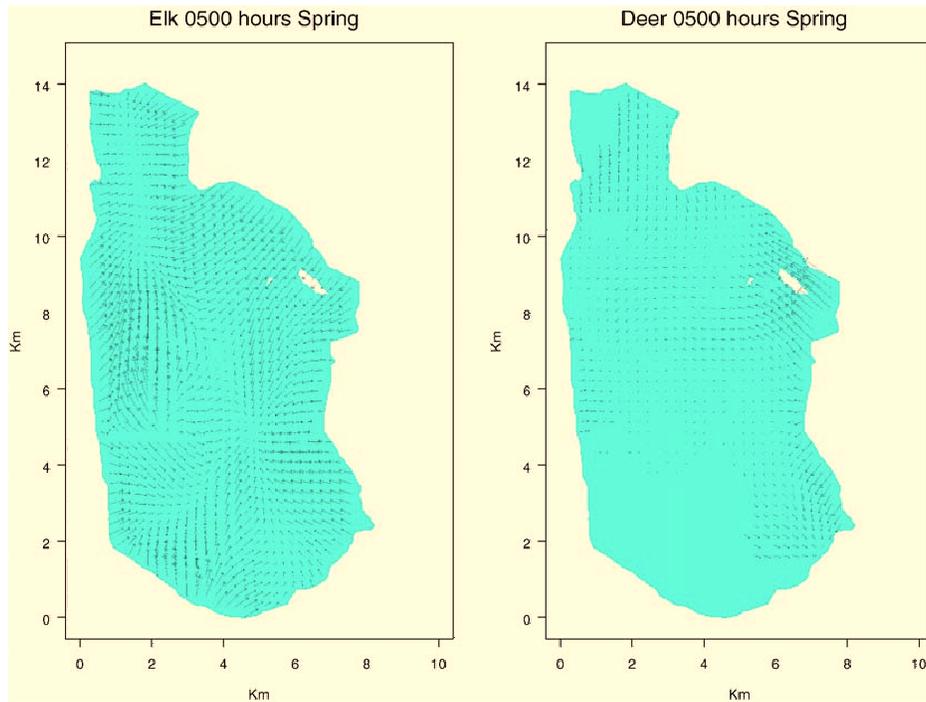


Fig. 7. Vector field estimates for the elk and deer.

7. Residual analysis

The use of residuals is an art where some physical scientists long maintained a significant lead on most, if not all, statisticians. ... Many of the more powerful forms of statistical analysis developed since 1920 can be formulated in terms of residuals. (Tukey, 1961)

The iterative and interactive interplay of summarizing by fit and exposing by residuals is vital to effective data analysis. (Tukey and Wilk, 1966)

The models need not fit perfectly or even adequately to prove usefully insightful. (Tukey and Wilk, 1966)

Fig. 8 provides the logarithm of the mod-squared residuals after fitting the function $\mu(\mathbf{r}, \langle t \rangle)$ to the elk data displayed as parallel boxplots against time of day. The results suggest that the 24 h rhythm needs to be present in the *noises* of model (5) in addition to in μ . One notes that the peaks in the plots occur near the same locations as those of Fig. 2. Hence, the model

$$\frac{\Delta \mathbf{r}}{\Delta t} = \mu(\mathbf{r}, \langle t \rangle) + \Sigma(\mathbf{r}, \langle t \rangle) \cdot \text{noise}$$

is next fit.

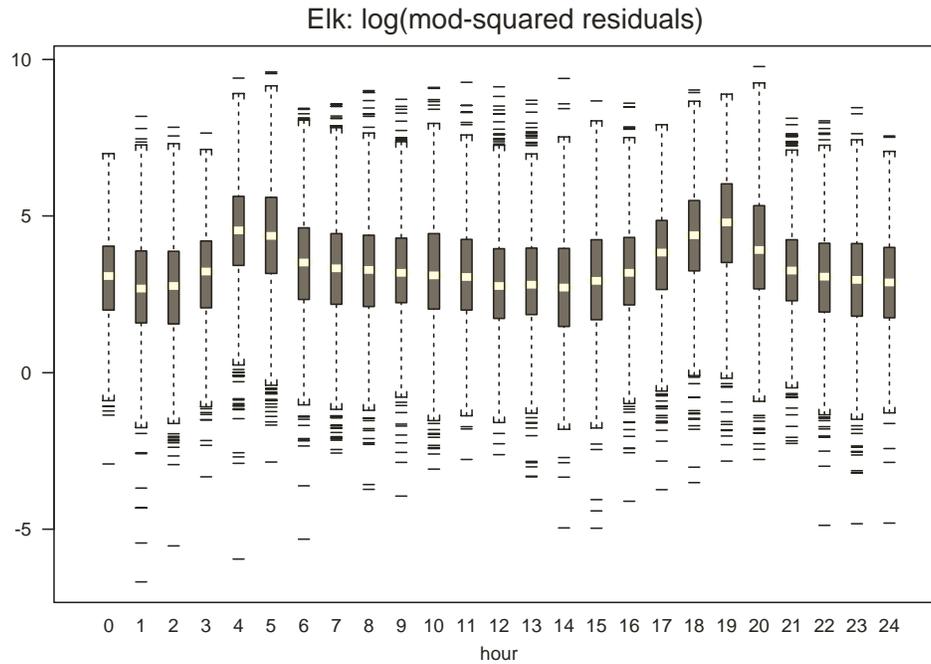


Fig. 8. Parallel boxplots of log(mod-squared of residuals) versus hour.

After the new fitting the residual spectral estimates for deer became consistent with an assumption of white noise so the figures are not presented. For the elk, Fig. 9 provides the analog of Fig. 5. The X -spectrum daily component is seen to be much reduced; however, the Y -component while reduced, remains. Initially this was puzzling. When attention turned to examination of the sampling times, t_{mk} , considered as a point process, it was found that they contained a daily variation. Fig. 10 provides their average periodogram.

To gain some understanding of this note that the spectral density matrix of a process \mathbf{V} sampled at the times of an independent stationary point process N is given by

$$c_N^2 \mathbf{f}_{\mathbf{V}\mathbf{V}}(\lambda) + \int f_{NN}(\lambda - \alpha) \mathbf{f}_{\mathbf{V}\mathbf{V}}(\alpha) d\alpha, \quad (6)$$

where c_N is the point process's rate and $f_{\mathbf{V}\mathbf{V}}$ is the spectral density matrix of the process \mathbf{V} , Brillinger (1984). If the process \mathbf{V} is nearly white, then the expression is nearly constant, see Appendix B. However, if it is not white the expression cannot be anticipated to be near constant. (The words “nearly” and “near” are used here because of the difficulties of defining white noise in continuous time.) Examining Fig. 9 this appears to be the case for the Y -component of the elk. There appears to be a fall-off as the frequency increases.

The residuals have led to a puzzle and Fourier analysis has led to its possible resolution, namely the driving process, \mathbf{V} , is non-Markov.

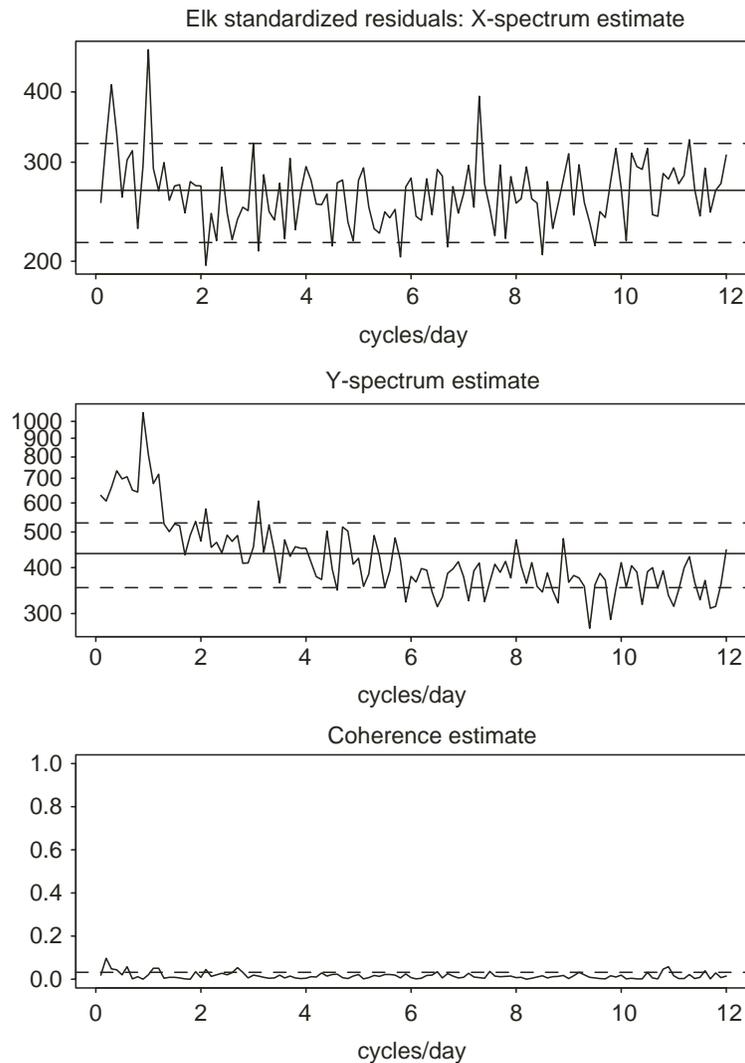


Fig. 9. Estimates of the residual spectrum of the X and Y coordinates and their coherence for the elk. The dashed lines provide approximate 95% bounds.

8. Dependencies amongst the animals?

To use FT does not imply that there are periodic phenomena. (There may be.)
(Tukey, 1980)

Consideration now turns to possible interactions amongst the animals. Perhaps the deer are meandering independently of each other. Perhaps the elk are doing so. Perhaps the deer are meandering independently of the elk. Fourier methods are employed to examine these possibilities for one elementary dependence model.

To simplify the description of the methods consider the complex-valued series, $U(t) = \dot{X}(t) + i\dot{Y}(t)$ for the deer and $V(t) = \dot{X}(t) + i\dot{Y}(t)$ for the elk, see Brillinger (1973). Next consider a time series analog of the equi-correlated variables model of

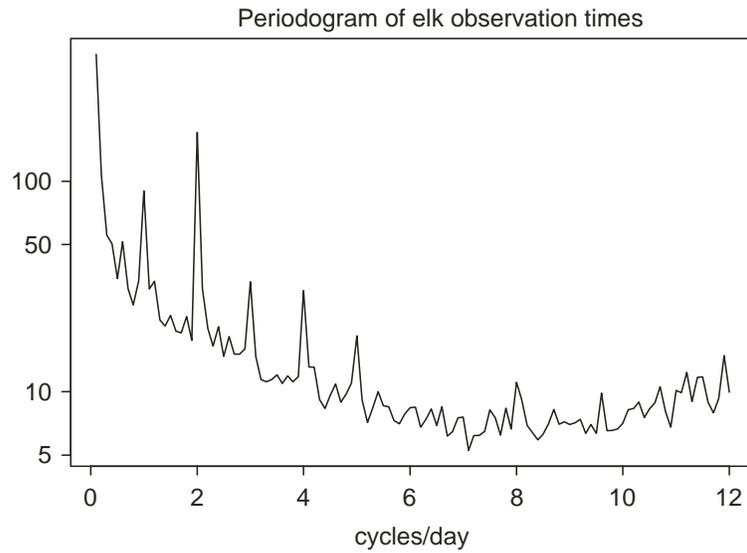


Fig. 10. The average of the periodograms of the sampling times.

multivariate analysis. Specifically suppose that, up to a complex conjugate, the coherency of the (complex-valued) velocity of any two of the elk (or deer) is the same. Further, assume that for any selected elk-deer pair the coherency is the same. There are thus three coherencies: R_{ee} for a pair of elk, R_{dd} for a pair of deer and R_{de} for an elk with a deer.

This structure would result were there independent stationary processes $\zeta(t), \varepsilon_e(t)$ such that

$$U_e(t) = \zeta(t) + \varepsilon_e(t) \quad (7)$$

for the elk, with e labeling the J elk. The random component process, ζ_e , may be viewed as representing time varying phenomena affecting all the elk simultaneously. It corresponds to the random effect models leading to equi-correlated variables in an analysis of variance context. It could include the effects of the day, e.g. a 24 h period.

A method of moments estimate of this coherency is

$$\hat{R}_{ee}(\lambda) = \left(\sum_{j \neq k} X_j \bar{X}_k \right) / \left((J-1) \sum |X_j|^2 \right), \quad (8)$$

where j and k sum over the J available elk and X_j denotes the approximate Fourier transform of the velocities, (3). This estimate is real valued. Following laws of large numbers and central limit theorems, when $\lambda \neq 0$ its approximate distribution for large J in the case that ζ is constant is that of

$$(|z|^2 - 1)/J, \quad (9)$$

where z is a standard complex normal variate, see Appendix B. This result may be used to assess the significance of $\hat{R}_{ee}(\lambda)$ and leads to the bounds in Figs. 11 and 12. Similar remarks may be made about $\hat{R}_d(\lambda)$ for the deer. The taper function in (2) deals with the holes that are present in the data.

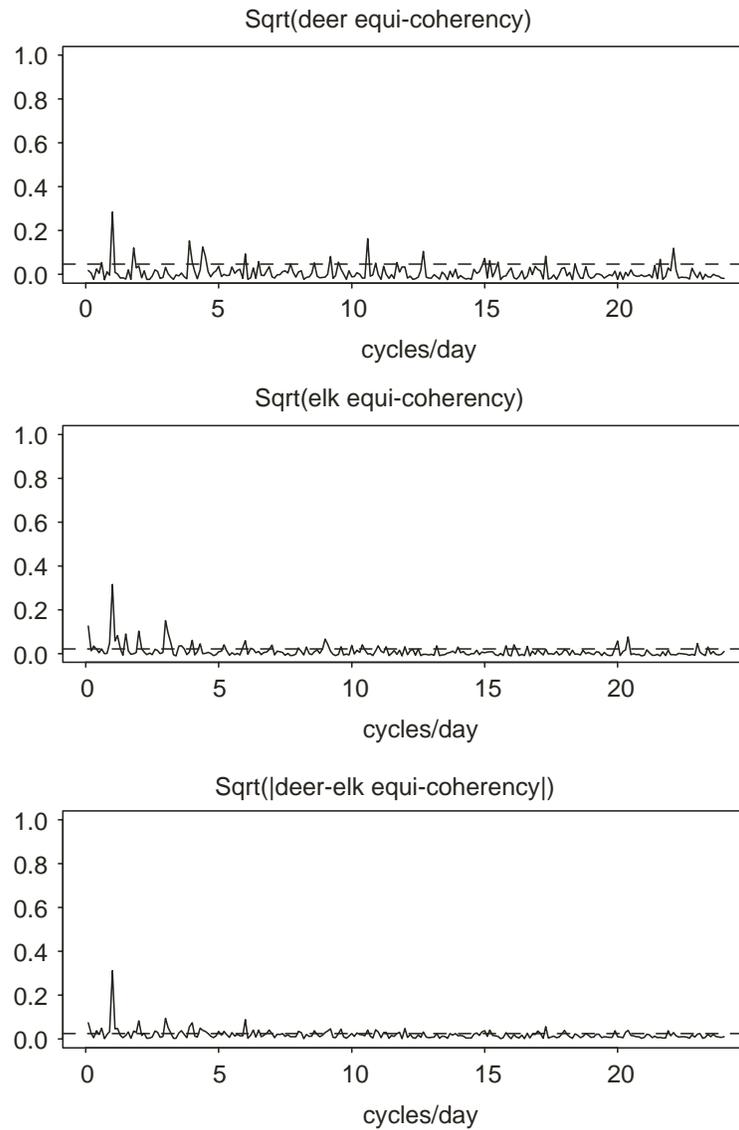


Fig. 11. Sqrt(abs(coherency)) statistics for deer–deer, elk–elk, deer–elk, respectively. The dashed lines correspond to approximate 95% upper null limits.

Turning to the coherency, $\hat{R}_{ed}(\lambda)$, between an elk and a deer a method of moments estimate is provided by

$$\hat{R}_{ed} = \left(\sum_{j,k} X_j \bar{Y}_k \right) / \sqrt{JK \sum |X_j|^2 \sum |Y_k|^2}, \tag{10}$$

where Y_k denotes the Fourier transform of the velocity of the k th deer. For large J, K the distribution of $|\hat{R}_{de}|$ may be approximated by that of a normal with mean $\pi / (4\sqrt{JK})$ and variance $(1 - \pi/4)/(JK)$, see Appendix B.

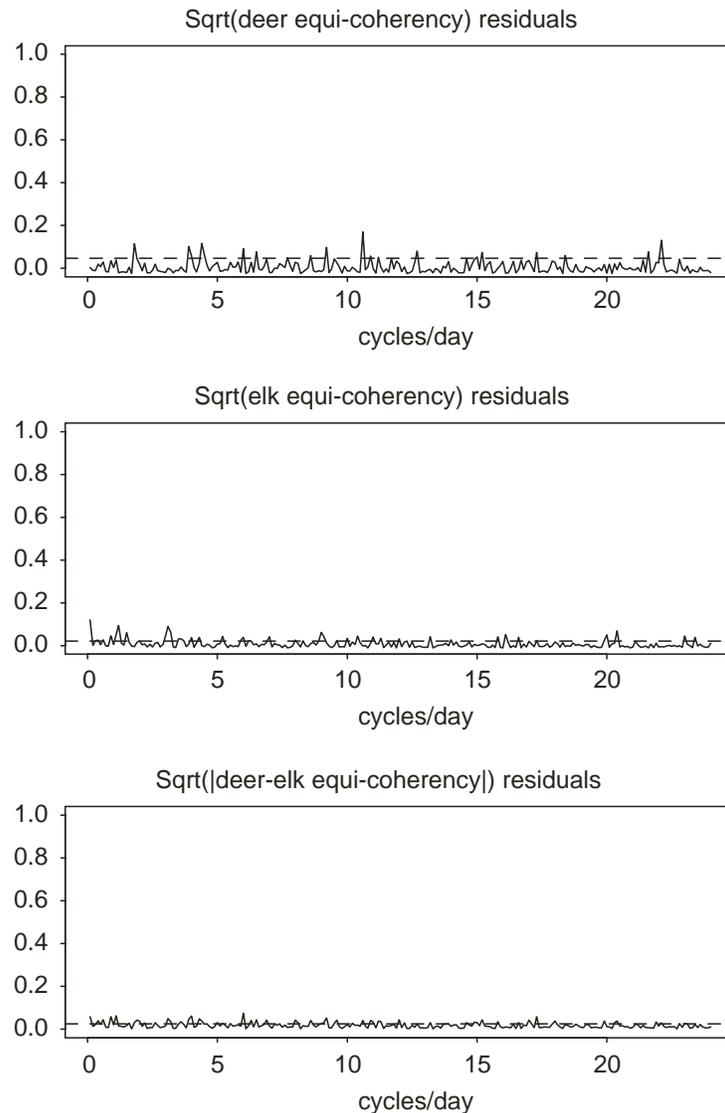


Fig. 12. $\text{Sqrt}(\text{abs}(\text{coherency}))$ statistics for deer, elk, deer–elk residuals, respectively.

Fig. 11 presents the estimates $\hat{R}_{dd}(\lambda)$, $\hat{R}_{ee}(\lambda)$, $|\hat{R}_{de}(\lambda)|$ based on the Fourier transform (3). The main structure apparent in the top two panels of the figure is the 24 h rhythm already noted in Figs. 2, 5, 6.

Fig. 12 provides the corresponding quantities based on the residuals i.e. the series having “removed” the 24 h period. Everything is flat and there is little evidence for a remaining latent component as assumed in model (7).

In summary, for the (contemporaneous hidden variable) model (7) no apparent dependence is found amongst the elk, or amongst the deer or between the elk and the deer beyond their common dependence on the time of day. This does not rule out other types of dependence. It might be that the deer lag behind the elk, as in $U(t) = V(t - \delta)$

for some δ , for example. Further it is reasonable to anticipate the strength of dependence between two animals depending on the distance between them. Such a model will be constructed in later work.

9. Discussion and conclusions

In summary, data analysis, like experimentation, must be considered as an open-ended, highly interactive, iterative process, ... (Tukey and Wilk, 1966)

A number of questions of interest were listed in the Introduction. The first two concerned providing descriptions of the animals' paths. This has been done to the extent that now realizations of the paths can be generated, for example to use in bootstrap computations. It appeared that the motion was not a Markov process, as it would have been were the driving process of (1) Brownian. A non-Markov driver surely had to be the case for there are irregular time lags in the animals' behavior, e.g. in respect of foraging bouts. Further an initial analysis was carried out to describe the interactions, if any, amongst the animals. None was found, but the model employed was limited.

The EDA attitude has proved a powerful one with which to investigate this data set of the paths of elk and deer in a large reserve. So too have the tools of smoothing and Fourier analysis. All three of these technologies were favorite techniques of John Tukey. In particular Fourier analysis has handled observations that were not made simultaneously, nor equi-spaced and with enforced gaps. These are typical characteristics of data obtained from animal telemetry studies. Smoothing played a central role in estimating drift patterns. In other work the results are related to the habitat of the Starkey area.

Acknowledgements

We thank the Editor and referees. Their comments were exceedingly helpful in improving the paper. The work of DRB was supported by the NSF Grants DMS-97-04739 and 99-71309.

Appendix A

The calculations made in Sections 5 and 8 use the Fourier transform of a function observed at unequally spaced times. JWT suggested an approximate method for such computations. It is presented below. Luckily modern computers are often so quick that the Fourier transform can often be evaluated directly from the definition.

There follows the text of a letter that JWT wrote to DRB January 22, 1980.
Dear David:

More to the cheaper calculation of

$$K(w) = \frac{1}{N} \sum_1^N e^{iwt_i}.$$

Let first

$$t_i = m_i h + \sigma_i$$

with m_i integral and $|\sigma_i| \leq n/2$. Then

$$K(w) = \frac{1}{N} \sum_1^N e^{iwm_i h} (1 + iw\sigma_i - w^2\sigma_i^2 + \dots).$$

Now put

$$k_0(m) = \frac{1}{N} \sum_{m_i+m} 1,$$

$$k_1(m) = \frac{1}{N} \sum_{m_i+1} \sigma_i,$$

$$k_2(m) = \frac{1}{N} \sum_{m_i+1} (\sigma_i^2/2)$$

...

so that

$$K(w) = \sum e^{iwmh} k_0(m) + \sum e^{iwmh} k_1(m) + \sum e^{iwmh} k_2(m) + \dots,$$

which can now be calculated by a few FFTs.

If we want to use w 's up to w_{MAX} and require to approximate $e^{iw\sigma}$ to 1%, then we can use the first two terms for

$$|w\sigma| \leq w_{\text{MAX}} \frac{h}{2} \leq 0.14 \approx \frac{1}{7}$$

that is with

$$h \leq \frac{1}{3.5w_{\text{MAX}}}$$

not too far from $(10)^{-1}(\pi/w_{\text{MAX}})$. To meet the same constraint with the first three terms would require

$$w_{\text{MAX}} \frac{h}{2} \leq 0.39$$

or

$$h \leq \frac{0.195}{w_{\text{MAX}}}.$$

The ratio of the numbers in the intervals is

$$\frac{0.39}{0.14} = 2.7,$$

which is greater than the ratio of numbers of FFTs, namely 1.5, so that going as far as the third term may easily be worthwhile.

Similar calculations give

| terms = FFTs | max $ w\sigma $ | ratio product |
|--------------|-----------------|---------------|
| 2 | 0.14 | 14 |
| 3 | 0.39 | 7 |
| 4 | 0.7 | 5.7 |
| 5 | 1 | 5 |

Suggesting that it may pay to go to 4 or 5 terms.

Regards, John

Appendix B

For elk j , write quantity (3) as X_j as in Section 8. Consider expression (8). It may be manipulated to

$$\left(\left(\left| \sum X_j \right|^2 \right) / \sum |X_j|^2 - 1 \right) / (J - 1).$$

Now when ζ in (7) is constant the X_j will be approximately independent and identically distributed. Laws of large numbers and central limit theorems will apply directly and one sees that the large sample distribution is that of (9) as indicated. The large sample distribution for (10) suggested in Section 8 follows similarly.

That (8) is a method of moments estimate of the coherency $R_{dd}(\lambda)$ may be argued as follows. Define a process $P(\cdot)$ via

$$dP(t) = V(t) dN(t)$$

with $V(\cdot)$ the complex-valued process representing the velocity of a deer and with $N = \{\tau_k\}$ a point process of measurement times. Now the empirical FT (3) is like

$$d_P^T(\lambda) = \int_0^T h(t) e^{-i\lambda t} dP(t). \tag{11}$$

The properties of such FTs are developed in Brillinger (1972, 1973). To proceed it will be assumed that the process $P(\cdot)$ is mixing and has stationary increments.

Supposing further that its real and imaginary parts are uncorrelated, but with the same autocovariance function, (Figs. 5 and 6 suggest this is not strongly contradicted), then for large T the distribution of (11) is approximately

$$N^C \left(0, 2\pi \int |H(\alpha)|^2 d\alpha f_{PP}(\lambda) \right)$$

with H the FT of h . The FTs, X_i, X_j , for two elk will be approximately complex normal with complex-valued correlation the coherency $R_{ee}(\lambda)$ leading to the proposed estimate (8).

Consideration next turns to the logic lying behind the model assessment via the periodogram of the residuals. The spectral density matrix of a sampled continuous time process was given as (6). Suppose now that the basic process is considered at closely spaced times $t = 0, \pm 1, \dots$, then the data may be represented as $\{N(t)V(t), t = 0, \pm 1, \pm 2, \dots\}$ where $\{V(t), t = 0, \pm 1, \pm 2, \dots\}$ is the series and $\{N(t), t = 0, \pm 1, \pm 2, \dots\}$ is a 0–1 valued series representing the sampling times, $\{t_k\}$. N takes on the value 1 when the V observation is present and 0 otherwise. The nonzero values of NV are the $\{V(\tau_k)\}$. The series NV reflects the statistical properties of the series V and N . For example if V and N are mutually independent and stationary with power spectra $f_{VV}(\lambda), f_{NN}(\lambda)$ then the process NV has power spectrum

$$c_N^2 f_{VV}(\lambda) + \int_{-\pi}^{\pi} f_{NN}(\lambda - \alpha) f_{VV}(\alpha) d\alpha$$

assuming further that V has mean 0 and that N has mean c_N . This expression was given in Bloomfield (1970) and follows from Example 2.10.4 in Brillinger (1975). From it one sees that if V is white noise then the spectrum of NV is constant. The heuristics of this result are clear: if the series is white then the sampled values are a separate selection and themselves white. One has a means of assessing whether a sampled time series is white. (These remarks correct some inappropriate discussion in Brillinger (2000).)

References

- Bloomfield, P., 1970. Spectral analysis with randomly missing observations. *J. Royal Statist. Soc.* 32, 369–380.
- Brillinger, D.R., 1972. The spectral analysis of stationary interval functions. In: Le Cam, L.M., Neyman, J., Scott, E.L. (Eds.), *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, pp. 483–513.
- Brillinger, D.R., 1973. An empirical investigation of the Chandler wobble and two proposed excitation processes. *Bull. Internat. Statist. Inst.* 45 (3), 413–434.
- Brillinger, D.R., 1975. *Time Series: Data Analysis and Theory*. Holt-Rinehart, New York.
- Brillinger, D.R., 1984. Statistical inference for irregularly observed processes. In: Parzen, E. (Ed.), *Time Series Analysis of Irregularly Observed Data*. Springer, New York, pp. 38–57.
- Brillinger, D.R., 2000. Estimating an irregularly sampled time series for whiteness. *Resenhas* 4, 423–431.
- Brillinger, D.R., Preisler, H.K., Ager, A.A., Kie, J., Stewart, B.S., 2001a. Modelling movements of free-ranging animals. Technical Report 610, Statistics Department, University of California, Berkeley.
- Brillinger, D.R., Preisler, H.K., Ager, A.A., Kie, J., 2001b. The use of potential functions in modelling animal movement. In: Saleh, M.H.K. (Ed.), *Data Analysis from Statistical Foundations*. Nova Science, New York, pp. 369–386.

- Brillinger, D.R., Preisler, H.K., Ager, A.A., Kie, J., Stewart, B.S., 2002. Employing stochastic differential equations to model wildlife motion. *Bull. Brazilian Math. Soc.* 33, 385–408.
- Cleveland, W.S., Grosse, E., Shyu, W.M., 1992. Local regression models. In: Chambers, J.M., Hastie, T.J. (Eds.), *Statistical Models in S*. Wadsworth, Pacific Grove, pp. 309–376.
- Dutt, A., Rokhlin, V., 1993. Fast Fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.* 14, 1368–1393.
- Hastie, T., 1992. Generalized additive models. In: Chambers, J.M., Hastie, T.J. (Eds.), *Statistical Models in S*. Wadsworth, Pacific Grove, pp. 249–308.
- Kloeden and Platen (1995). *Numerical Solution of Stochastic Differential Equations*. Springer, New York.
- Mallows, C.L., Tukey, J.W., 1982. An overview of techniques of data analysis, emphasizing its exploratory aspects. In: Tiago de Oliveira, J., Epstein, B. (Eds.), *Some Recent Advances in Statistics*. Academic Press, London.
- Preisler, H.K., Brillinger, D.R., Ager, A.A., Kie, J., 1999. Analysis of animal movement using telemetry and GIS data. *Proceedings of the ASA Section on Statistics and the Environment*, Washington, DC, pp. 100–105.
- Preisler, H.K., Brillinger, D.R., Ager, A.A., Kie, J., 2001. Stochastic differential equations: a tool for studying animal movement. In: Rennolls, K. (Ed.), *Proceedings of IUFRO4.11 Conference: ‘Forest Biometry, Modelling and Information Science’*, University of Greenwich, June 25–29, 2001, 23pp. <http://cms1.gre.ac.uk/conferences/iufro/proceedings/>.
- Rowland, M.M., Bryant, L.D., Johnson, B.K., Noyes, J.H., Wisdom, M.J., Thomas, J.W., 1997. The Starkey project: history, facilities, and data collection methods for ungulate research. US Department of Agriculture, Forest Service, Pacific Research Station, General Technical Report PNW-GTR-396:1-62.
- Tukey, J.W., 1961. Data analysis and behavioral science or learning to bear the quantitative man’s burden by shunning badmandements. In: Jones, L.V. (Ed.), *The Collected Works of John W. Tukey IV*. Wadsworth, Monterey, pp. 187–389.
- Tukey, J.W., 1971. First Scott lecture. In: Brillinger, D.R. (Ed.), *The Collected Works of John W. Tukey II*. Wadsworth, Monterey, pp. 857–884.
- Tukey, J.W., 1973. Exploratory data analysis: as part of a larger whole. *Proceedings of the 18th Conference on Design of Experiments in Army Research and Development I*, Washington, DC, pp. 1–10.
- Tukey, J.W., 1975. Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians, Vancouver, BC, Vol. 2. Canadian Mathematical Congress, Montreal, Canada*, pp. 523–531.
- Tukey, J.W., 1977. *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
- Tukey, J.W., 1979. Comment on Parzen’s “Nonparametric statistical modelling”. *J. Amer. Statist. Assoc.* 74, 121–122.
- Tukey, J.W., 1980. Personal communication. Overheads of talk at National Center for Atmospheric Research.
- Tukey, J.W., 1984. *The Collected Works of John W. Tukey I*. Wadsworth, Monterey.
- Tukey, J.W., Wilk, M.B., 1966. Data analysis and statistics: an expository overview. In: *29 AFIPS Conference Proceedings*, Montvale, NJ. pp. 695–709.