Hidden Markov models for circular and linear-circular time series

Hajo Holzmann[1], Axel Munk[1], Max Suster[2] and Walter Zucchini[1]

[1]Georg-August-Universität and [2]McGill Centre for Research in Neuroscience

Abstract

We introduce a new class of circular time series based on hidden Markov models. These are compared with existing models, their properties are outlined and issues relating to parameter estimation are discussed. The new models conveniently describe multi-modal circular time series as dependent mixtures of circular distributions. Two examples from biology and meteorology are used to illustrate the theory. Finally we introduce a hidden Markov model for bivariate linear-circular time series and use it to describe larval movement of the fly Drosophila.

Keywords: animal behaviour; circular correlation; circular data; mixtures; von Mises distribution; wind direction; wrapped distributions

1 Introduction

The goal of this paper is to describe a new class of models for circular-valued time series, namely circular hidden Markov models (HMMs). We outline their main properties and illustrate their application.

Directional or circular data, i.e. measurements in the form of angles or other periodic values, arise naturally in several scientific disciplines such as biology, climatology, oceanography, geophysics and astronomy. The substantial literature on the analysis of circular data focuses on regression and correlation analysis, and on one- and multisample tests. (See e.g. Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001).) However, circular data with a temporal structure, i.e. circular-valued time series, have not received as much attention. Typical examples of such time series are hourly or daily wind directions at a fixed location, the series of directions followed by an animal over time, or the time of day at which some event occurs, e.g. the level of air pollutant peaks.

Apart from summary statistics and graphical displays it is generally of interest to describe observations by means of a stochastic model that encapsulates their characteristic features. The latter include of course any temporal dependence that is evident in the series. Thus an important concept here is the circular autocorrelation function (cacf) introduced by Fisher and Lee (1994). (See Appendix A for the definition.) HMMs offer considerable flexibility in their serial dependence properties.

In many applications the marginal distribution of the observations is clearly multimodal, suggesting that they arose from a mixture of different distributions associated with different regimes. Precisely such behaviour is a key property of HMMs; their marginal distribution is a mixture of state-dependent distributions which can differ in their means, dispersions, or which may even belong to different distributional families. This degree of flexibility is not achieved by the models that have been described to date.

Bivariate time series in which one of the variables is linear-valued and the other circular-valued also arise in practice, e.g. measurements on the speed and the direction of wind, ocean current, or an animal movement taken at consecutive points in time; the daily peak load of a pollutant and the time of day at which the peak is observed. We are not aware of any general-purpose models that are suitable for such time series. We show that HMMs provide tractable models for linear-circular time series and illustrate their use to describe the movements of Drosophila larvae.

The remainder of the paper is organised as follows. Following a brief review of the main classes of existing models for circular time series (Section 2) we define circular HMMs and consider two varieties of such models, the von Mises and the wrapped normal or Cauchy HMMs (Section 3). An expression is given for the likelihood and some issues relating to parameter estimation are discussed, including the findings based on a simulation experiment. In Section 3 we give two examples of application. We show that a von Mises-HMM provides a plausible description of the changes in direction of Drosophila larval movements. The marginal distribution can be modeled by a mixture of two von Mises distributions with equal means and unequal dispersion parameters. The second example deals with a series of wind directions. The raw data were not given as continuous measurements; they had been classified into the 16 conventional compass directions, i.e. as N, NNE, NE, ..., NWN. We show how the likelihood of the HMMs can be easily modified to take precise account of such censoring. This demonstrates that the models can be applied to discrete-valued circular time series. Section 4 outlines the construction of HMMs for bivariate linear-circular times series. We discuss the issue of measuring the linear-circular cross-correlation structure for such series and illustrate the application of the model using a series of bivariate observations on the speed and change in direction of Drosophila larvae. Section 5 gives some conclusions and some general remarks. The Appendices A and B give the required theoretical details on circular autocorrelation- and cross-correlation functions.

All data used in this article are available at *http://www.statlab.de/statistics_group/*.

2 Outline of existing models

We outline a number of models that have been proposed in the literature for time series of circular-valued observations.

Linked processes

Fisher and Lee (1994) introduced a general method to construct circular processes from linear processes. The idea is to associate (link) the circular-valued observations with those of a linear-valued time series, for which models are plentiful.

A *link function* is a strictly monotonically increasing, odd function $g : \mathbb{R} \to (-\pi, \pi)$. Recall that a circular random variable (crv) is by definition a random variable taking values on the unit circle $\mathbb{R}/2\pi\mathbb{Z}$. If X is a linear random variable and $\mu \in [0, 2\pi)$, then $\Theta = g(X) + \mu \bmod 2\pi$ is called the linked crv. Conversely, if Θ is a crv having a density, then $X := g^{-1} \circ (\Theta - \mu)$, where $\Theta - \mu$ is taken in $(-\pi, \pi)$, is a linear random variable. Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary process on the real line and $\mu \in [0, 2\pi)$. The process $\Theta_t = g(X_t) + \mu \bmod 2\pi$ is called the *linked circular process*. It is also stationary. Important examples are obtained by taking $(X_t)_{t\in\mathbb{Z}}$ to be ARMA(p,q).

Linked processes are designed for time series with low dispersion. In series with high dispersion the occurance of similar values of Θ_t does not imply that the corresponding X_t are also similar; they can differ substantially. Thus linked processes are unlikely to be useful for such series.

The cacf of these processes is not available in closed form, but simulations indicate that it is similar to the acf of the corresponding linear process (Fisher and Lee, 1994).

Circular autoregressive processes

Fisher and Lee (1994) also introduced a von Mises-based analog of the well-known AR(p) processes for linear data. The density of a von Mises distribution $VM(\mu, \kappa)$ is given by

$$\phi(\theta) = [2\pi I_0(\kappa)]^{-1} \cdot \exp[\kappa \cos(\theta - \mu)], \quad \kappa \geq 0,\ \mu \in [0, 2\pi).$$

where

$$I_n(\kappa) := [2\pi]^{-1} \int_0^{2\pi} \exp(\kappa \cos\phi) \cdot \cos(n\phi)\, d\phi$$

is the modified Bessel function of order $n \geq 0$.

A process $(\Theta_t)_{t\in\mathbb{Z}}$ is called *circular autoregressive* of order p, CAR(p), with link function g if Θ_t, given $\Theta_{t-1} = \theta_{t-1}, \Theta_{t-2} = \theta_{t-2}, \cdots, \Theta_1 = \theta_1$, is distributed as VM($\mu$,$\kappa$) for $t = 1, \cdots, p$, and as VM(μ_t,κ) for $t > p$, where

$$\mu_t := \mu + g\left(\alpha_1 \cdot g^{-1}(\theta_{t-1} - \mu) + \cdots + \alpha_p \cdot g^{-1}(\theta_{t-p} - \mu)\right),$$

and where $\theta_i - \mu$ is taken in $(-\pi, \pi)$. Again the cacf can only be obtained by simulation.

Wrapped processes

A simple way to obtain a crv is to wrap a linear random variable around the circle, i.e. to reduce its values mod 2π. If applied to a linear process $(X_t)_{t\in\mathbb{Z}}$, the resulting circular process $\Theta_t := X_t \bmod 2\pi$ is called the corresponding *wrapped circular process*. A particular case occurs if X_t is an AR(p) process, in which case the wrapped process is called a *wrapped autoregressive process* WAR(p). These were introduced by Breckling (1989) (see also Fisher and Lee (1994) for further discussion). The cacf is given in Appendix A.

Markov processes

Johnsson and Wehrly (1979) introduced the following method to obtain stationary Markov processes on the circle. Let $f,\ g$ be two densities on the circle; set $F(\theta) = \int_0^\theta f(\eta)\, d\eta$ and

$$p(\theta|\eta) = 2\pi g\big(2\pi[F(\theta) - F(\eta)]\big) f(\theta).$$

For convenience we assume that $\theta,\ \eta \in [0, 2\pi)$. For each η, $p(\cdot|\eta)$ is a density on the circle and $\int_0^{2\pi} p(\theta|\eta) \cdot f(\eta)\, d\eta = f(\theta)$. Hence we obtain a stationary Markov process $(\Theta_t)_{t\geq 0}$ such that the joint density of $\Theta_0, \ldots, \Theta_T$ is given by

$$f(\theta_0, \ldots, \theta_T) = f(\theta_0) \cdot \prod_{t=1}^{T} p(\theta_t | \theta_{t-1}).$$

Notice that for uniform g, $(\Theta_t)_{t \geq 0}$ is i.i.d. with density f. If g is von Mises $VM(\mu, \kappa)$ and f is uniform, then

$$p(\theta | \eta) = [2\pi \cdot I_0(\kappa)]^{-1} \cdot \exp[\kappa \cos(\theta - \eta - \mu)].$$

Johnsson and Wehrly (1979) used this model to develop uniformly most powerful tests for independence in a circular time series. They did not give the cacf, but this can be derived by direct calculation and is given in Appendix A.

3 Hidden Markov models for circular time series

Hidden Markov models were introduced by L. E. Baum and his co-workers in a series of papers published between 1966 and 1972 (e.g. Baum and Petrie (1966); for additional references to these papers see e.g. Chapter 2 of MacDonald and Zucchini (1997)). There is now an extensive literature on the subject covering both theory and a wide variety of applications. Among other things, properties of the maximum likelihood estimators such as strong consistency (Leroux, 1992) and asymptotic normality (Bickel, Ritov and Rydén, 1998) have been obtained. Their application to discrete-valued and multivariate time series have been discussed in detail in MacDonald and Zucchini (1997). However, to our knowledge, their use as general-purpose models for circular-valued time series has not yet been described.

3.1 Hidden Markov models

We start by defining circular HMMs in general and then consider two particular families of models, the von Mises-HMM and the wrapped-normal-HMM.

Let $(C_t)_{t \geq 1}$ be an irreducible, homogeneous Markov chain with state space $\{1, \ldots, m\}$ and transition probability matrix $\mathbf{\Gamma} := (\gamma_{ij})_{i,j=1,\ldots,m}$. By the irreducibility of (C_t), there exists a unique, strictly positive, stationary distribution $\mathbf{\Delta} = (\delta_1, \ldots, \delta_m)$. We shall suppose that (C_t) is stationary, so that Δ is for all t the distribution of C_t, and $\mathbf{\Delta} \cdot \mathbf{\Gamma} = \mathbf{\Delta}$. The powers of $\mathbf{\Gamma}$ are denoted as $\mathbf{\Gamma}^k = (\gamma_{ij}^{(k)})_{i,j=1,\ldots,m}$.

A circular process $(\Theta_t)_{t \geq 1}$ will be called a circular HMM with underlying stationary Markov chain $(C_t)_{t \geq 1}$ if for every $T \geq 1$, $i_1, \ldots, i_T \in \{1, .., m\}$ and $A_1, \ldots, A_T \subset [0, 2\pi)$ measurable we have

$$P(\Theta_1 \in A_1, \ldots, \Theta_T \in A_T \,|\, C_1 = i_1, \ldots, C_T = i_T) = \prod_{t=1}^{T} P(\Theta_t \in A_t \,|\, C_t = i_t).$$

It follows that for $t \leq T$,

$$P(\Theta_t \in A_t \,|\, C_1 = i_1, \ldots, C_T = i_T) = P(\Theta_t \in A_t \,|\, C_t = i_t).$$

Hence Θ_{t_0} depends on (C_t) only through C_{t_0}. If the state dependent probabilities $P(\Theta_t \in A_t \,|\, C_t = i)$ do not depend on t, which we will always assume, the process $(\Theta_t)_{t\geq 1}$ is also stationary.

We note that it is possible to relax the assumption that the process (C_t) is stationary, or the assumption that state dependent probabilities do not depend on t. Such extensions are useful, and necessary, if one wishes to model series that exhibit trend or seasonal variation, or series that depend on covariates other than time. Examples of non-stationary HMMs are given in MacDonald and Zucchini (1997).

The von Mises-HMM

The von Mises-HMM is defined by assuming the state dependent probabilities $P(\Theta_t \in \cdot | C_t = i)$ to be distributed as $VM(\mu_i, \kappa_i),\ i = 1, 2, \ldots, m$. The cacf of a von Mises HMM is given in Appendix A. Of particular interest are those cases in which either all mean directions, or all concentration parameters, are assumed to be equal, i.e. $\mu_1 = \ldots = \mu_m$, or $\kappa_1 = \ldots = \kappa_m$.

The wrapped normal- and wrapped Cauchy-HMM

The wrapped normal distribution, $WN(\mu, \rho)$, is obtained by wrapping a normal distribution $N(\mu, \sigma)$ around the circle, where $\rho = e^{-\sigma^2/2}$. Specifically, if $\phi(x; \mu, \sigma)$ is the density of the normal distribution, then the density of $WN(\mu, \rho)$ is given by (cf. Jammalamadaka and SenGupta (2001), p. 44)

$$\phi_w(\theta; \mu, \rho) = \sum_{n\in\mathbb{Z}} \phi(\theta + 2\pi n; \mu, \sigma) \ .$$

The wrapped normal-HMM is defined by assuming that the state dependent probabilities $P(\Theta_t \in \cdot | C_t = i)$ are distributed as $WN(\mu_i, \rho_i),\ i = 1, 2, \ldots, m$. Again one can consider cases of equal mean directions, $\mu_1 = \ldots = \mu_m$, or equal mean resultant lengths, $\rho_1 = \ldots = \rho_m$. The wrapped Cauchy distribution $WC(\mu, \rho)$ is defined in an analogues way; one simply assumes that the state dependent probabilities are distributed as wrapped Cauchy. The cacfs of the wrapped normal- and the wrapped Cauchy-HMM are given in Appendix A.

As was indicated by Fisher and Lee (1994), the cacf does not seem to be well suited for discriminating between different classes of models, e.g. between circular-HMM and CAR(p), but it may be useful to select an appropriate model within a given class. For the class of HMMs the cacf can give some indication of the number of hidden states (See Appendix A). If the mean directions of the state-dependent distributions are all equal then the cacf is 0 for all lags ≥ 1 and thus the cacf is of no use for selecting m in this case. However, as is the case in the application described in Section 4.1, if the empirical cacf has this form then an HMM with equal mean directions would be plausible. Note that this lack of circular correlation does not imply that the consecutive values are independent.

3.2 Maximum likelihood estimation

In this section we will discuss maximum likelihood estimation in circular HMMs. To obtain the likelihood function, suppose that the conditional distributions $P(\cdot | C_t = i)$ have densities $\phi_{\eta_i}(\theta)$, where the η_i range over some parameter set $\mathcal{E}$. Then the joint density of $\Theta_1, ..., \Theta_T$

conditional on $C_1 = i_1, ..., C_T = i_T$ is given by

$$f_T(\theta_1, \ldots, \theta_T \,| C_1 = i_1, \ldots, C_T = i_T) = \prod_{t=1}^{T} \phi_{\eta_{i_t}}(\theta_t),$$

hence the joint density of $\Theta_1, \ldots, \Theta_T$ is

$$\begin{aligned}
f_T(\theta_1, \ldots, \theta_T) &= \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} f_T(\theta_1, \ldots, \theta_T | C_1 = i_1, \ldots, C_T = i_T) \\
&\quad \cdot P(C_1 = i_1, \ldots, C_T = i_T) \\
&= \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} \delta_{i_1} \phi_{\eta_{i_1}}(\theta_1) \prod_{t=2}^{T} \gamma_{i_t, i_{t+1}} \cdot \phi_{\eta_{i_t}}(\theta_t) \\
&= \mathbf{\Delta} \cdot \mathbf{A}(\theta_1) \cdot \mathbf{\Gamma} \cdot \mathbf{A}(\theta_2) \cdot \ldots \cdot \mathbf{\Gamma} \cdot \mathbf{A}(\theta_T) \cdot \mathbb{1}
\end{aligned}$$

where $\mathbf{A}(\theta) = \text{ diag } (\phi_1(\theta), \ldots, \phi_m(\theta))$ and $\mathbb{1} = (1, ..., 1)'$. If we let

$$\mathbf{B}(\theta_t) = \mathbf{\Gamma} \cdot \mathbf{A}(\theta_t),$$

then using $\mathbf{\Delta} \cdot \mathbf{\Gamma} = \mathbf{\Delta}$ we can also write

$$f_T(\theta_1, \ldots, \theta_T) = \mathbf{\Delta} \cdot \prod_{t=1}^{T} B_t(\theta_t) \cdot \mathbb{1}. \tag{1}$$

The parameters over which (1) is maximized are $\eta_1, \ldots, \eta_m$ and the entries of $\mathbf{\Gamma}$, where the latter are subject to constraints; in particular each row sum is equal to one.

The von Mises-HMM (continued)

For the case $m = 2$ of the von Mises-HMM, the parameters are $0 < \gamma_{12}, \gamma_{21} < 1$, $\mu_1, \mu_2 \in [0, 2\pi)$ and $0 < \kappa_1, \kappa_2 < \infty$. The marginal distributions of Θ_t are mixtures of von Mises distributions. The question of identifiability needs to be considered for consistent estimation of the parameters (Leroux, 1992). Identifiability has been proved by Fraser, Hsu and Walker (1981). If we assume that $\kappa_1 = \kappa_2 = \kappa$ and restrict this parameter to a compact interval $[\epsilon, 1/\epsilon]$ for some small $\epsilon > 0$, then the conditions for strong consistency and asymptotic normality of the maximum likelihood estimator (MLE) are satisfied (Leroux, 1992; Bickel *et al.*, 1998). In case of different concentration parameters the likelihood function becomes unbounded, a problem that is well known from the theory of mixture models. (See Spurr and Koutbeiy (1991) and Hathaway (1985) for simulations and consistency results.) Although consistency results have not yet been obtained for HMMs, our simulations show that maximum likelihood estimation seems to work well in this situation, too.

The wrapped normal- and wrapped Cauchy-HMM (continued)

Similar qualifications regarding consistency apply to the wrapped normal and wrapped Cauchy HMM. In case of two states ($m = 2$) the parameter ρ has to lie in some interval $[\epsilon, 1 - \epsilon]$ for small $\epsilon > 0$. Identifiability of finite mixtures was obtained by Holzmann, Munk and Stratmann (2003).

The MLE can be computed by direct maximization of the likelihood function, e.g. using the Newton-Raphson algorithm. In order to avoid underflow, a scaled version of the likelihood and finally the log-likelihood can be calculated as suggested by MacDonald and Zucchini (1997). To this end consider the vectors of forward probabilities defined by

$$\alpha_j = \boldsymbol{\Delta} \cdot \prod_{t=1}^{j} \mathbf{B}(\theta_t), \quad j = 1, \ldots, T.$$

Of course, $\alpha_{j+1} = \alpha_j \cdot \mathbf{B}(\theta_{j+1})$. In each step we scale the vector α_j to have mean 1, i.e. compute

$$s_1 = \sum_{i=1}^{m} \alpha_1(i), \quad \tilde{\alpha}_1 = s_1^{-1} \cdot \alpha_1,$$

and recursively

$$\alpha'_{j+1} = \tilde{\alpha}_j \cdot \mathbf{B}(\theta_{j+1}), \; s_{j+1} = \sum_{i=1}^{m} \alpha'_{j+1}(i), \; \tilde{\alpha}_{j+1} = \frac{1}{s_{j+1}} \cdot \alpha'_{j+1}.$$

Then $\tilde{\alpha}_T = \alpha_T/(s_1 \cdot \ldots \cdot s_t)$, $f_T(\theta_1, \ldots, \theta_T) = \sum \alpha_T(i)$ and $\sum \tilde{\alpha}_T(i) = 1$, hence

$$\log(f_T(\theta_1, \ldots, \theta_T)) = \sum_{t=1}^{T} \log(s_t).$$

In carrying out the numerical maximization one must also attend to the problem that some of the parameters are subject to constraints. A simple option here is to reparameterize the model in terms of parameters that are unconstrained. For example, in the case $m = 2$, the entries $0 < \gamma_{12} < 1$, $0 < \gamma_{21} < 1$ of $\boldsymbol{\Gamma}$ can be reparameterized in terms of their logits. The logit transformation can also be used for the parameters ρ in wrapped normal, and the log transformation for the parameters κ in the von Mises. The case of HMMs with more than two states is a little trickier but appropriate parameterizations which lead to unconstrained maximization are suggested in Zucchini and MacDonald (1998).

The practical problem of finding starting values requires close examination of the data and some experimentation. For the mean directions the modes of the histogram provide obvious starting values so long as the modes are clearly distinguishable. Of course more sophisticated methods such as nonparametric mode estimators are also available. For the concentration parameters, we suggest partitioning the sample into two parts at one antimode and estimating κ for each subsample separately. The starting values for the parameters of Γ seem to have little effect on the convergence properties so long as the entries are not close to 0 or 1.

The case where the modes are not clearly distinguishable is trickier. Some suggestions are offered below. Desirable here would be a procedure that yields automatic (i.e. purely data-based) starting values, but this is not yet available. Furthermore it seems unlikely that a single method will work well for all models.

3.3 Simulation results

In this section we summarize the results of a Monte Carlo study which we conducted to assess the performance of the estimation procedure for a number of two-state von Mises-

HMMs. Random deviates from the von Mises distribution were generated using an algorithm by Best and Fisher (1979).

First we simulated series with parameters $\mu_1 = 1,\ \mu_2 = 5,\ \kappa_1 = 3,\ \kappa_2 = 8,\ \gamma_{1,2} = 0.2,\ \gamma_{1,2} = 0.3$. The sample size was chosen as 100 and 500, respectively, and 200 replicates were performed in each simulation scenario. This case is clearly bimodal and so presents no starting-value problems.

The resulting estimates are displayed as boxplots in Figures 1 and 2. These show that the estimators of all six parameters are approximately unbiased. A QQ-plot revealed that the μ-estimates are approximately normal even for $n = 100$. The distributions of the κ- and the γ-estimates are skewed for $n = 100$ but symmetric (and approximately normal) for $n = 500$. This gives an indication that a fairly long series is required for asymptotic normality to take effect for these parameters.

Figures 1 and 2

In a second experiment we considered equal mean directions with $\mu = 3$ and $\kappa_1 = 1,\ \kappa_2 = 10,\ \gamma_{1,2} = 0.2,\ \gamma_{1,2} = 0.1$. To obtain starting values we first treated the observations as if they were a random sample from a *single* von Mises distribution and estimated μ and κ accordingly. We then set $\kappa_1 = \kappa/2,\ \kappa_2 = 2\kappa$. The results, displayed in Figure 3, show similar tendencies as in the first experiment but the skewness is slightly more extreme. (Five outlining estimates of κ_2 are not shown in the plot.)

Figure 3 and 4

Finally, we considered the case of equal concentration parameters with $\kappa = 4$ and $\mu_1 = 2,\ \mu_2 = 3,\ \gamma_{1,2} = 0.7,\ \gamma_{1,2} = 0.8$. Here the resulting marginal distribution is unimodal. We note also that it is for this case ($\kappa_1 = \kappa_2$) that the asymptotic normality of the estimators has been proved. (It is not unlikely that this also holds for the other cases considered, but this has not yet been proved.)

Starting values were again obtained by first estimating μ and κ as one would if the observations were a random sample from a single von Mises distribution; μ_1 was set equal to μ plus the circular standard deviation, and μ_1 as μ minus the standard deviation. The results are displayed in Figure 4.

For $n = 100$ there appears to be a small bias in $\gamma_{2,1}$. This is probably due to the few cases where the estimate converged to 1. Apart from this the distributions the μ- and γ-estimators are approximately normally distributed in accordance with the asymptotic theory for $n = 100$. However, the estimator of κ is still skew even for $n = 500$. As a practical remedy one can reparameterize the von Mises distribution in terms of $\log(\kappa)$, whose estimator is approximately normally distributed even for $n = 100$.

4 Examples

In this section we describe two examples of application of circular HMMs.

4.1 Changes in direction of Drosophila larval movement

Larvae of the fruit fly Drosophila melanogaster disperse by repeated episodes of linear movement and brief episodes of head swinging and turning (Green, Burnet and Conolly, 1983). Locomotion can be largely summarized by the speed and direction change of the larva in each of these two episodic states (Suster et al., 2003). During straight movement, larvae maintain a high speed and a low direction change, in contrast to the low speed and high direction change characteristic of turning episodes. An accurate mathematical model of Drosophila larval crawling behavior would be valuable for characterizing subtle changes in locomotion due to genetically-targeted alterations in the underlying neural circuitry. Previously, randomly selected pairs of speed and direction change were used to model this motor behavior (Suster, 2000). Using this approach, we could not accurately simulate the pattern of larval locomotion, indicating the need to take time and variable correlations into account. Given that the larva alternates between one of two states during locomotion (linear movement or turning), we reasoned that a two-state HMM, in which both speed and turning rate are modeled according to two behavioral states, would be more appropriate for describing the pattern of larval locomotion. As an illustration we will examine the movements of a single larva, whose position was recorded once per second over three minutes. The path taken by the larva is displayed in Figure 5.

Figures 5 and 6

From these data it is easy to compute the speed and the change in direction in each second. In this section we will focus on the circular variable, i.e. the time series of direction change. Most of the directions are concentrated near 0, but there are also several observations at higher angles. Thus the marginal distribution seems to be a mixture of two (or more) circular distributions, suggesting that a two-state circular HMM might be appropriate. State 1, which is more frequently visited, involves small changes in direction and hence a high concentration parameter. In state 2, the data is spread over the whole circle; the concentration parameter is small. Taking into account the cacf of the series which has no significant values for lags ≥ 1, we selected a common mean direction for both states (in which case cacf = 0 for lags ≥ 1). The estimated parameters for a von Mises-HMM are $\hat{\mu} = -0.01$, $\hat{\kappa}_1 = 36.38$, $\hat{\kappa}_2 = 1.73$ and $\hat{\gamma}_{1,2} = 0.24$, $\hat{\gamma}_{2,1} = 0.34$. This gives a log-likelihood of -114 and a stationary initial distribution $\hat{\delta}_1 = 0.59$, $\hat{\delta}_2 = 0.41$. Figure 6 shows a histogram of the data together with the estimated marginal mixture distribution and its weighted components. Finally we tested for independence of the observations within our class of models using the likelihood ratio test (Giudici, Rydén and Vandekerhove, 2000). In case of two states, the i.i.d. situation arises if $\gamma_1 = 1 - \gamma_2$. Under the hypothesis of independence, the likelihood ratio test statistic has a chi–squared distribution with one degree of freedom. For the above larva, the value of the likelihood ratio statistic is 10.8, hence we can reject the hypothesis of independence.

We will return to this example in Section 5 to fit a bivariate linear-circular HMM which describes the joint behaviour of speed and changes of direction of the larvae.

4.2 Wind direction at Koeberg, South Africa

In this example we consider a time series comprising average hourly values of the wind direction at the Koeberg nuclear power station in South Africa. The period covered is 1 May 1985 to 30 April 1989 which results in a series of length 35064. The averages were classified into the 16 conventional directions N, NNE, ..., NNW, coded 1 to 16, in that order. MacDonald and Zucchini (1997) fitted a number of models for this series, including a multinomial–HMM with two states and a 16 state Markov chain. A complication here is that the distribution of wind direction changes over the day, a feature that is clearly evident in the rose-plots displayed in Figure 7.

Figure 7

They also considered a reduced series of daily data for a particular hour of the day. Their models did not take into account the fact that the observations arose from circular-valued measurements, but regarded the 16 directions simply as 16 categories. Here we will also consider the daily series of the average wind direction in the interval 23:00-24:00, and fit the von Mises-HMM to model the observations.

The data in this application are not available as exact measurements; they are given as principal compass directions, with 16 possible values. It is therefore necessary to modify the likelihood function of the von Mises-HMM so that it takes account of this interval-censoring. It is remarkable that HMMs can accommodate this type of complication; the exact likelihood is easy to obtain, and to compute. We discretize the von Mises distribution by integrating the density over arcs of equal length around the discretized observations. Thus the probability that the wind direction is in the principle compass direction j $(j = 1, \ldots 16)$ is given by

$$\pi_j = \int_{l_j}^{u_j} \phi(t; \mu, \kappa)\, dt, \; l_j := \frac{j - 0.5}{16} \cdot 2\pi, \; u_j := \frac{j + 0.5}{16} \cdot 2\pi,$$

where ϕ is the density of a von Mises distribution with parameters μ and κ. Thus, for each $i = 1, 2, \ldots, m$, the $i\,th$ state-dependent probability distribution for the discretized observations is discrete, namely multinomial with parameters $(\pi_1^i, \pi_2^i, \ldots, \pi_{16}^i)$, where the π_j^i are determined by μ_i and κ_i, the parameters of the corresponding von Mises state-dependent distribution. The expression for the likelihood for the discretized data is obtained using these m multinomial distributions in place of the corresponding m state-dependent von Mises densities in (1).

We consider the issue of model selection which, in the case of a von Mises-HMM, comes down to selecting m, the number of states in the Markov chain. Figure 8 displays the fitted marginal distributions (mixtures of Mises distributions) for $m = 1, 2, 3, 4$ plotted over the histogram of the observed directions.

Figure 8

Clearly the case $m = 1$ (a single von Mises distribution) does not provide an acceptable fit for these data, which exhibit two clear modes in the directions SSE and NNW. As is to be expected, the fit improves with increasing m, but it is not clear at which stage the improvement might be a consequence of overfitting. A more objective assessment is provided by the values of formal model selection criteria. The table below which gives, for $m = 1, 2, 3, 4$, the number of free parameters, k, the negative log likelihood, $-l$, and the values of the AIC and BIC criteria.

model	k	-log(llk)	AIC	BIC
1-state von Mises-HMM	2	3932.0	7868.1	7878.7
2-state von Mises-HMM	6	3606.7	7225.4	7257.1
3-state von Mises-HMM	12	3538.3	7100.5	7163.9
4-state von Mises-HMM	20	3485.4	7010.9	7116.6
2-state multinomial-HMM	32	3461.9	6987.8	7156.9
Saturated Markov chain	240	3292.5	7065.0	8333.9

The criteria are also given for two additional models that were fitted in MacDonald and Zucchini (1997). The first is a two-state multinomial-HMM, in which the directions are regarded as 16 unrelated categories. (Note that although the discretized von Mises-HMM also leads to multinomial state-dependent distributions, the parameters of these distributions are *related* because each set of 16 parameters is uniquely determined by the two parameters of the corresponding von Mises distribution.) The second model is a saturated stationary Markov chain, whose 16×16 transition probability matrix is therefore determined by $16 \times 15 = 240$ free parameters.

The BIC rates the four-state von Mises-HMMs as best in the group; the AIC rates the 2-state multinomial-MMM, but it too indicates that (at least) $m = 4$ states should be used in the von Mises-HMM. The parameter estimates for that model are given by

$$\widehat{\Gamma} = \begin{pmatrix} 0.642 & 0.020 & 0.000 & 0.338 \\ 0.143 & 0.268 & 0.589 & 0.000 \\ 0.000 & 0.328 & 0.669 & 0.003 \\ 0.719 & 0.000 & 0.041 & 0.240 \end{pmatrix} \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \\ \hat{\mu}_4 \end{pmatrix} = \begin{pmatrix} 2.72 \\ 6.01 \\ 1.90 \\ 5.12 \end{pmatrix} \begin{pmatrix} \hat{\kappa}_1 \\ \hat{\kappa}_2 \\ \hat{\kappa}_3 \\ \hat{\kappa}_4 \end{pmatrix} = \begin{pmatrix} 8.49 \\ 12.95 \\ 0.82 \\ 1.50 \end{pmatrix}.$$

While it is somewhat disappointing to select a model with so many parameters it needs to be kept in mind that the series is a long one. Furthermore at least two of the four state-dependent densities (Figure 9) correspond to interpretable meteorological states. State 1, the most likely of the four states, is strongly associated with the direction SSE and state 2 (even more strongly) with the direction NNW. The prevailing directions in states 3 and 4 are slightly east of those in states 1 and 2, respectively.

Figure 9

The above conclusion needs to be qualified by the fact, illustrated in Figure 10, that wind direction is highly seasonal over the year. It is possible to relax the assumption that the state process is stationary, and the assumption that the state-dependent distribution depends solely on the state of the process; it can be allowed to depend on time, or some other covariate. Among other things such extensions allow one to accommodate seasonal variation. A discussion on the issue of seasonality would lead us beyond the scope of this paper. (Examples and applications of seasonal HMMs are given in, e.g., MacDonald and Zucchini, 1997.)

Figure 10

5 Hidden Markov models for bivariate linear-circular time series

We start by defining bivariate linear-circular HMMs. Let $(C_t)_{t\geq 1}$ be an irreducible stationary Markov chain. A bivariate process $(X_t, \Theta_t)_{t\geq 1}$, where the Θ_t are circular and the X_t linear random variables, respectively, is called a linear-circular HMM with underlying Markov chain $(C_t)_{t\geq 1}$ if for all $T \in \mathbb{N}$, $A_t \subset [0, 2\pi)$, $B_t \subset \mathbb{R}$ measurable and $i_t \in \{1, \dots, m\}$,

$$P(X_1 \in B_1, \dots, X_T \in B_T,\ \Theta_1 \in A_1, \dots, \Theta_T \in A_T \,|\, C_1 = i_1, ..., C_T = i_T) = \prod_{t=1}^{T} P(X_t \in B_t, \Theta_t \in A_t \,|\, C_t = i_t)$$

Modeling is simplified considerably if one assumes conditional independence, i.e.

$$P(\Theta_t \in A_t, X_t \in B_t \,|\, C_t = i_t) = P(\Theta_t \in A_t \,|\, C_t = i_t) \cdot P(X_t \in B_t \,|\, C_t = i_t).$$

In Appendix B we suggest a way to quantify the cross-correlation structure of such linear-circular HMMs.

The main advantage of the above assumption is that it allows us to model the (state-dependent) linear and the circular random variables individually rather than jointly. In other words we can use univariate rather than bivariate state-dependent distributions. However, either with or without the assumption, the likelihood of linear-circular HMMs is still of the form given in (1). The state–dependent distributions now pertain to two random variables, but so long as one can compute values of their joint density functions (or probability functions) no additional complications arise. We therefore proceed immediately to an example of application.

A bivariate model for speed and change in direction of Drosophila larvae

The time series of the spatial positions of Drosophila larvae at consecutive equispaced times were transformed to provide the (more conveniently interpretable) bivariate linear-circular time series of speeds and changes in direction, which we will refer to as angles.

An examination of the speeds and the angles for these larvae (Figure 11 gives a graphical display for one larva) indicates that high speeds are associated with small angles and low speeds with large angles. These two modes of behaviour suggest the existence of two states, one being fast directed movement and the other slow undirected movement. Furthermore the states appear to be somewhat persistent.

Figure 11

Thus these series appear to exhibit the properties of a two-state linear-circular HMM in which one state is associated with a high mean for the linear variable (speed) and a low variance for the circular variable (angle). The situation is reversed in state 2. As our model for these time series we have used a conditionally independent HMM with von Mises state-dependent distributions for the angles and gamma distributions for the speeds. The density of the Gamma distribution $G(\nu, \lambda)$ is given by

$$f(x) = [\lambda^{\nu} \cdot \mathbf{\Gamma}(\nu)]^{-1} \cdot x^{\nu-1} \exp(-x/\lambda) \cdot \mathbb{1}_{[0,\infty)}(x), \quad \lambda, \nu > 0.$$

For the numerical maximization of the likelihood we reparameterized the gamma distribution in terms of its expectation, ξ, and variance σ^2, (where $\xi = \lambda \cdot \nu$, $\sigma^2 = \lambda^2 \cdot \nu$, so $\nu = \xi^2/\sigma^2$, $\lambda = \sigma^2/\xi$) because this leads to improved numerical performance.

The estimated parameters for the above mentioned larva are

$$\hat{\mu} = -0.02, \quad \begin{array}{llll} \hat{\kappa}_1 = 31.76, & \hat{\xi}_1 = 1.19, & \hat{\sigma}_1^2 = 0.1, & \hat{\gamma}_{1,2} = 0.24 \\ \hat{\kappa}_2 = 1.81, & \hat{\xi}_2 = 0.45, & \hat{\sigma}_2^2 = 0.05, & \hat{\gamma}_{2,1} = 0.32 \end{array}$$

This gives a log-likelihood of -169 and a stationary distribution $\hat{\delta}_1 = 0.57$, $\hat{\delta}_2 = 0.43$. Notice that as we guessed, in state 1 both concentration parameter and mean velocity are higher than in state 2.

Figure 12 shows a plot of the sample cross-correlation function of the series for one larva and the theoretical cross-correlation function for the estimated parameter values. For the first two lags the fit is very close but the empirical cacf remains higher than expected under the model. To accommodate this feature of the data would require a substantially more complex model.

Figure 12

6 Conclusions

We believe that the HMMs described in this paper provide a versitile class of general-purpose models for circular-valued time series. It was illustrated how easy it is to extend such models to describe bivariate linear-circular time series. In a similar manner one can adapt them to describe multivariate circular time series, such as wind direction, or the times of day at which some event (e.g. peak demand for electricity) occurs at different locations.

Many of their probabilistic properties, such as the marginal and conditional distributions, the moment functions and the forecast distribution are easy to derive. As was illustrated in the application in Section 5 they can be modified to cope with interval-censoring in a precise manner. Missing observations, so long as these are missing "completely at random", present no difficulties and can be dealt with (again precisely) by adding a few lines of code to the software.

Despite the simplicity and flexibility of HMMs our knowledge of their properties is far from complete and many statistical questions remain to be answered. Included here is a lack of results on the properties of the parameter estimators, their standard errors and the construction of accurate confidence intervals. Additional techniques for diagnostic checking need to be developed. Model selection needs to be addressed in greater detail, and in particular methods to cope with the selection bias that arises when one uses the same data to select a model, fit the selected model and also to assess the fit.

One of the attractive features of HMMs is the fact that their likelihood is easy to compute, but often enough it is not an easy function to maximize. In most non-trivial applications the likelihood of an HMM has multiple local maxima and so convergence of the maximization algorithm provides no guarantee that the global maximum has been found. This applies not only to the direct maximization of the likelihood but also to the well-known Baum-Welch algorithm (the EM approach to parameter estimation for HMMs) that we did not discuss in this paper.

We also wish to draw attention to the growing literature on the Bayesian analysis of HMMs, in particular on the issue of model selection, i.e. deciding how many states, m, should be used in the Markov chain component of the model. (For key references to this literature see, e.g., Scott, 2002.) From a Bayesian point m is one of the unknown parameters to which one should assign a prior and then compute it's posterior distribution. The amount of computation required (by means of Monte Carlo Markov Chains) is formidable, an order of magnitude greater than is needed to compute the BIC, which is an asymptotic approximation to the Bayes factors under a uniform prior.

References

Baum, L.E. and Petrie, T. (1966) Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*, **37**, 1554-1563.

Best, D. and Fisher, N. (1979) Efficient simulation of the von Mises distribution. *Applied Statistics*, **24**, 1527.

Bickel, P. J., Ritov, Y. and Rydén, T. (1998) Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *The Annals of Statistics*, **26**, 1614-35.

Breckling, J. (1989) *The Analysis of Directional Time Series: Applications to Wind Speed and Direction.* Springer Verlag.

Fisher, N. I. and Lee, A. J. (1994) Time series analysis of circular data. *Journal of the Royal Statistical Society Series B*, **56**, 327-39.

Fisher, N. I. and Lee, A. J. (1983) A correlation coefficient for circular data. *Biometrika*, **70**, 327-32.

Fraser, M. D., Hsu, Y. S. and Walker, J. J. (1981) Identifiability of finite mixtures of von Mises distributions. *Annals of Statistics* **9** 1130-1131.

Green, C. H., Burnet, B. and Conolly, K. J. (1983) Organization and patterns of inter- and intraspecific variation in the behaviour of Drosophila larvae. *Animal Behaviour*, **31**, 282-91.

Giudici, P., Rydén, T. and Vandekerhove, P. (2000) Likelihood-ratio test for hidden Markov models. *Biometrics*, **56**, 742-7.

Hathaway, R.J. (1985) A constrained formulation of maximum-likelihood estimation for normal mixtures distributions. *The Annals of Statistics*, **13**, 795-800.

Holzmann, H., Munk, A. and Stratmann, B. (2003) Identifiability of finite mixtures-with applications to circular distributions. preprint.

Jammalamadaka, S. R. and SenGupta, A. S. (2001) *Topics in Circular Statistics.* World Scientific Publishing Co.

Johnson, R. A. and Wehrly, T.E. (1979) Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, **66**, 255-6.

MacDonald, I.L. and Zucchini, W. (1997) *Hidden Markov and other Models for Discrete-valued Time Series* . Chapman & Hall.

Mardia, K.V. (1976), Linear-circular correlation coefficients and rhythmometry. *Biometrika*, **63**, 403-5.

Mardia, K.V. and Jupp, P. (2000) *Directional Statistics.* Wiley & Sons.

Scott, S.L. (2002) Baysian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, **97**, 337-351.

Spurr, B. D. and Koutbeiy, M.A. (1991) A comparision of various methods for estimating the parameters in mixtures of von Mises distributions. *Communications in statistics / Simulation and Computation*, **20**, 725-41.

Suster, M. L. (2000) Neural control of larval locomotion in Drosophila melanogaster. PhD Thesis, University of Cambridge.

Suster, M. L., Martin, J. R., Sung, C. and Robinow, S. (2003) Targeted expression of tetanus toxin reveals sets of neurons involved in larval locomotion in Drosophila. *Journal of Neurobiology* **55**, 233-46.

Zucchini, W. and MacDonald, I.L. (1998) Hidden Markov time series models: some computational issues. *Computing Science and Statistics*, ed. S. Weisberg, Interface Foundation of North America, **30**, 157-63.

Appendix A

Let $(\Theta_t)_{t\geq 0}$ be a stationary circular process. The *circular autocorrelation function* (cacf) is defined by

$$\rho_C(k) := \rho_C(\Theta_0, \Theta_k), \quad k \geq 0,$$

where ρ_C denotes the circular correlation coefficient as introduced by Fisher and Lee (1983). By stationarity we can write

$$\rho_C(k) = \frac{E[\cos\Theta_0 \cos\Theta_k] \cdot E[\sin\Theta_0 \sin\Theta_k] - E[\sin\Theta_0 \cos\Theta_k] \cdot E[\cos\Theta_0 \sin\Theta_k]}{(1 - E[\cos^2\Theta_0]) \cdot E[\cos^2\Theta_0] - (E[\sin\Theta_0 \cos\Theta_0])^2} \tag{2}$$

For properties of ρ_C, an estimator and its asymptotics see Fisher and Lee (1983).

Wrapped autoregressive processes

Let ρ_j denote the lag-j correlation of X_t and let $\sigma^2, \phi_1 \ldots \phi_p$ be its AR(p) parameters. Then

$$\rho_C(k) = \frac{\sinh\left(2\rho_k\sigma^2 \cdot [1 - \phi_1\rho_1 - \ldots - \phi_p\rho_p]^{-1}\right)}{\sinh\left(2\sigma^2 \cdot [1 - \phi_1\rho_1 - \ldots - \phi_p\rho_p]^{-1}\right)}$$

See Fisher and Lee (1983, 1994) for the calculation. The cacf is similar in form to the acf of the underlying AR(p)-process.

Markov processes

A direct calculation yields the cacf in the example of Section 1.

$$\rho_C(k) = \left[\cos^2(k\mu) - \sin^2(k\mu)\right] \cdot (A_1(\kappa))^{2k}.$$

Hidden Markov models

In order to compute the expectations in (2) and to determine the correlation structure of a circular HMM, first observe that if f, g are bounded measurable functions of period 2π and $k \geq 1$,

$$\begin{aligned} E(f(\Theta_0) \cdot g(\Theta_0)) &= \sum_{i=1}^{m} \delta_i \cdot E(f(\Theta_0) \cdot g(\Theta_0) \,|\, C_0 = i), \qquad (3) \\ E(f(\Theta_0) \cdot g(\Theta_k)) &= \sum_{i,j=1}^{m} \delta_i \cdot \gamma_{ij}^{(k)} \cdot E(f(\Theta_0) \,|\, C_0 = i) \cdot E(g(\Theta_k) \,|\, C_k = j). \end{aligned}$$

Here $E(\cdot|\, C_t = i)$ denotes expectation with respect to the conditional distribution $C_t = i$. In general, if

$$f_\mu(\theta) = (2\pi)^{-1}\big[a_0 + 2\sum_{p\geq 1} a_p \cos p\,(\theta-\mu)\big]$$

is the Fourier expansion of f_μ which is assumed to converge uniformly, then

$$\begin{aligned}\int_0^{2\pi} \sin x \cos x\, f_\mu(x)\,dx &= a_2 \sin(2\mu)/2,\\ \int_0^{2\pi} \cos^2 x\, f_\mu(x)\,dx &= [a_0 + a_2\cos(2\mu)]/2.\end{aligned}$$

For the von Mises HMM, $a_n = A_n(\kappa) := I_n(\kappa) \setminus I_0(\kappa)$. Since

$$E(\cos(\Theta_t)\,|\,C_t = i) = A_1(\kappa_i)\cos(\mu_i) \text{ and } E(\sin(\Theta_t)\,|\,C_t = i) = A_1(\kappa_i)\sin(\mu_i)$$

where $\phi_i(x) = [2\pi I_0(\kappa_i)]^{-1}\exp[\kappa_i\cos(\theta-\mu_i)]$ is the conditional density given $C_t = i$, using (3) the cacf in (2) can now be calculated. Note that if $\mu_1 = \ldots = \mu_m$, the correlation is 0 for all lags ≥ 1.

Figures 13 and 14

The Fourier expansion of the density $\phi_w(\theta;\mu,\rho)$ of the wrapped normal distribution $WN(\mu,\rho)$ is

$$\phi_w(\theta;\mu,\rho) = (2\pi)^{-1}\big[1 + 2\sum_{p\geq 1}\rho^{p^2}\cos p(\theta-\mu)\big] \tag{4}$$

hence $a_p = \rho^{p^2}$. Similarly, the density of the wrapped Cauchy distribution $WC(\mu,\rho)$ has Fourier expansion

$$\phi_w(\theta;\mu,\rho) = (2\pi)^{-1}\big[1 + 2\sum_{p\geq 1}\rho^{p}\cos p(\theta-\mu)\big], \tag{5}$$

hence $a_p = \rho^p$.

Appendix B

In order to analyze the cross-correlation structure of linear-circular time series, we need a measure for the correlation between a linear and a circular random variable. There have been several attempts to define such a quantity in the literature, see e.g. Mardia (2000, pp. 245-248). We will use the one introduced by Mardia (1976).
For a two-dimensional random vector (X,Y) let

$$\rho(X,Y) = \frac{E[XY] - EXEY}{\sqrt{(E[X^2]-[EX]^2)(E[Y^2]-[EY]^2)}}$$

denote its correlation. Let (X,Θ) be a linear-circular random vector. In order to determine whether there is a relation of the kind

$$X = a + b\cdot\cos(\Theta) + c\cdot\sin(\Theta),$$

for some constants $a, b, c \in \mathbb{R}$, Mardia (1976) considered the multiple correlation of X with $(\cos(\Theta), \sin(\Theta))$ defined by

$$R_{X,\Theta} :=$$
$$\left[\frac{\rho(X, \cos\Theta)^2 + \rho(X, \sin\Theta)^2 - 2 \cdot \rho(X, \cos\Theta)\rho(X, \sin\Theta)\rho(\cos\Theta, \sin\Theta)}{1 - \rho(\cos\Theta, \sin\Theta)^2}\right]^{1/2}.$$

The sample counterpart is given by replacing the correlations of the random variables by the sample correlation coefficients. The cross correlation function of a stationary linear-circular process is then defined by

$$c\rho_C(k) := \begin{cases} R_{X_k,\Theta_0} & : \quad k \geq 0 \\ R_{X_0,\Theta_{|k|}} & : \quad k < 0 \end{cases} \tag{6}$$

Hidden Markov models

The quantities which are needed to compute the cross-correlations of a linear-circular HMM are obtained similarly as in case of a circular HMM. For $k \geq 0$ we have that

$$E(X_k \cos\Theta_0) = \sum_{i,j=1}^{m} \delta_i \cdot \gamma_{ij}^{(k)} \cdot E(\cos\Theta_0 | C_0 = i) \cdot E(X_k | C_k = j),$$

and a similar expression holds for $E(X_k \sin\Theta_0)$ and for $k < 0$. As an example, consider the linear-circular HMM of Section 4. Then, using conditional independence,

$$\begin{aligned} E(X_0 \cos\Theta_0) &= \sum_{i=1}^{m} \delta_i \cdot (\lambda_i \nu_i) \cdot [\cos(\mu_i) A_1(\kappa_i)], & (7) \\ E(X_k \cos\Theta_0) &= \sum_{i,j=1}^{m} \delta_i \cdot \gamma_{ij}^{(k)} \cdot (\lambda_j \nu_j) \cdot [\cos(\mu_i) A_1(\kappa_i)], \quad k \geq 1, \\ E(X_0 \cos\Theta_{|k|}) &= \sum_{i,j=1}^{m} \delta_i \cdot \gamma_{ij}^{(|k|)} \cdot (\lambda_i \nu_i) \cdot [\cos(\mu_j) A_1(\kappa_j)], \quad k \leq -1. \end{aligned}$$
$$(8)$$

Using (3) and (7) the cross correlation function (6) can now be written down explicitely. Notice that if the Markov chain X_t is reversible, the cross correlation function is symmetric in the sense $c\rho_C(k) = c\rho_C(-k)$. This is always true for a stationary, two-state chain.

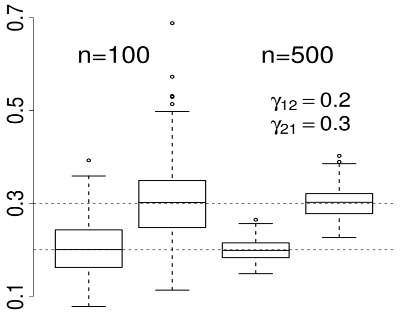

Figure 1: Boxplots of estimated transition probabilities for von Mises HMM with $(\mu_1, \mu_2, \kappa_1, \kappa_2, \gamma_{1,2}, \gamma_{1,2}) = (1, 5, 3, 8, 0.2, 0.3)$

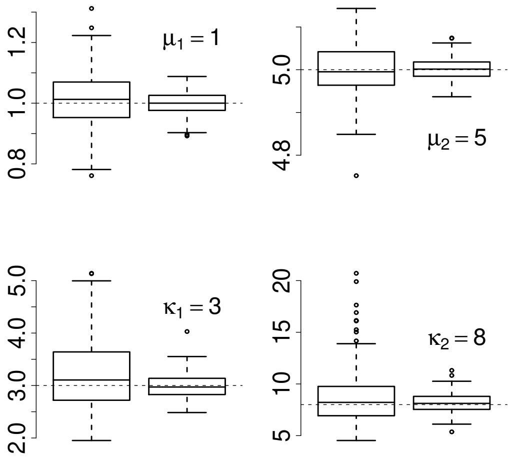

Figure 2: Boxplots of the estimated parameters of the state dependent distributions for von Mises HMM with $(\mu_1, \mu_2, \kappa_1, \kappa_2, \gamma_{1,2}, \gamma_{1,2}) = (1, 5, 3, 8, 0.2, 0.3)$

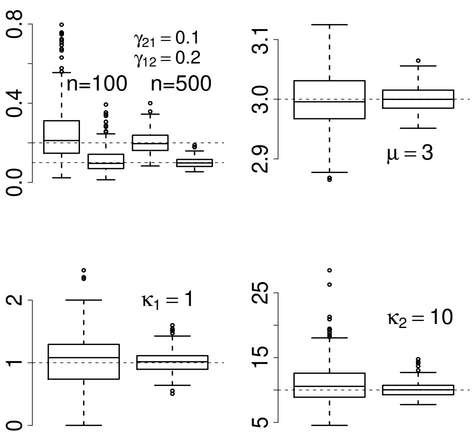

Figure 3: Boxplots of the estimated parameters for von Mises HMM with equal mean directions with $(\mu, \kappa_1, \kappa_2, \gamma_{1,2}, \gamma_{1,2}) = (3, 1, 10, 0.1, 0.2)$

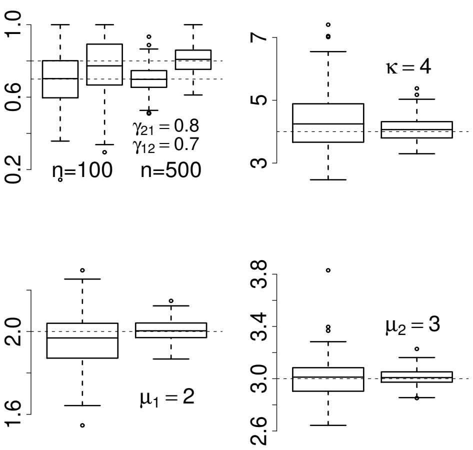

Figure 4: Boxplots of the estimated parameters for von Mises HMM with equal concentration parameter with $(\mu_1, \mu_2, \kappa, \gamma_{1,2}, \gamma_{1,2}) = (2, 3, 4, 0.8, 0.7)$

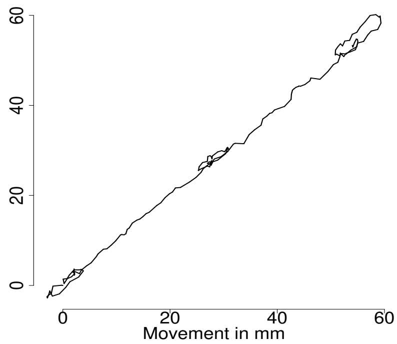

Figure 5: Plot of the movement of one larva of the fly Drosophila

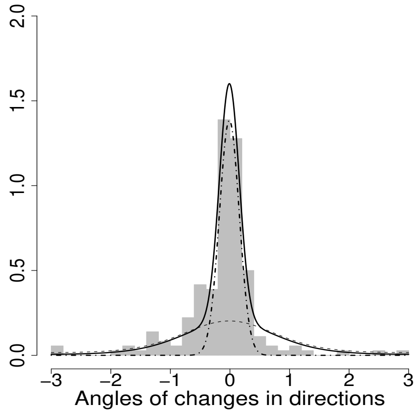

Figure 6: Histogram and fitted density (solid line) with weighted mixture components (dashed line: highly variable component, dashed dotted line: low variable component) of the ts of one Drosophila larva

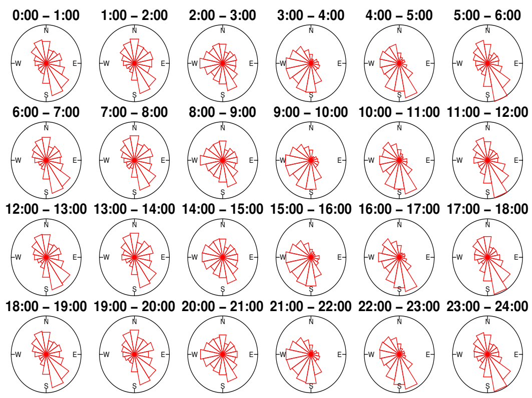

Figure 7: Rose plots of wind directions at Koeberg by time of day

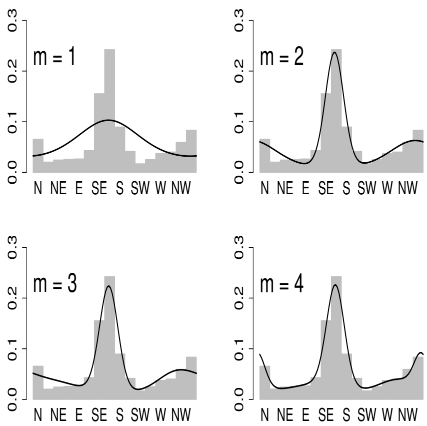

Figure 8: Histogram and fitted densities of the modified von Mises HMM with m hidden states for the ts of wind directions at Koeberg

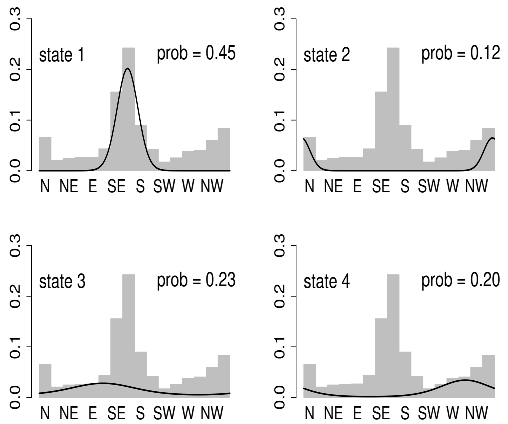

Figure 9: Histogram and fitted state-dependend densities in the modified von Mises HMM with $m = 4$ states for ts of wind directions at Koeberg

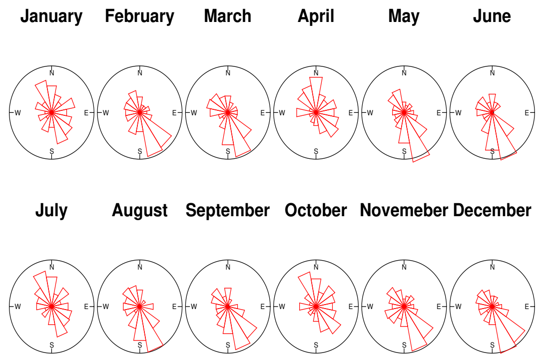

Figure 10: Rose plots of monthly wind directions (23:00-24:00)

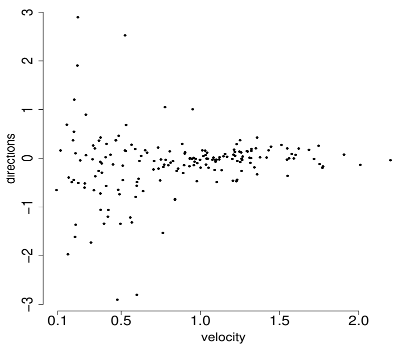

Figure 11: Joint plot of velocities and changes in directions of the movement of a Drosophila larva

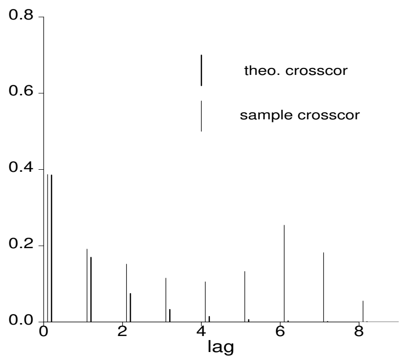

Figure 12: Linear- circular cross correlations for one larva with estimated parameters $(\mu, \kappa_1, \kappa_2) = (-0.02, 1.81, 31.76)$, $(m_1, m_2, \sigma_1^2, \sigma_2^2) = (0.45, 1.18, 0.05, 0.1)$, $(\gamma_{1,2}, \gamma_{2,1}) = (0.32, 0.24)$

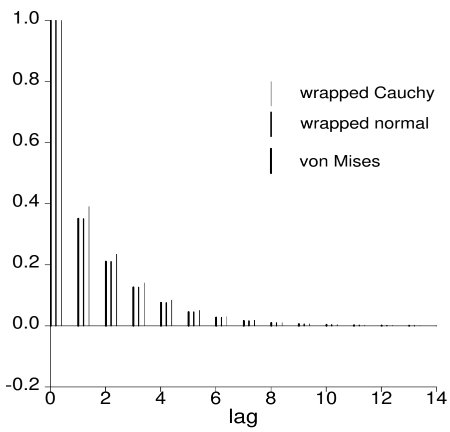

Figure 13: Cacf in several circular HMMs, parameter values $(\mu_1, \mu_2) = (1, 5)$, $\gamma_{1,2} = \gamma_{2,1} = 0.2)$, $\kappa_1 = \kappa_2 = 8$ (von Mises HMM) and $\rho_1 = \rho_2 = 0.94$ (wrapped Cauchy and normal HMM)

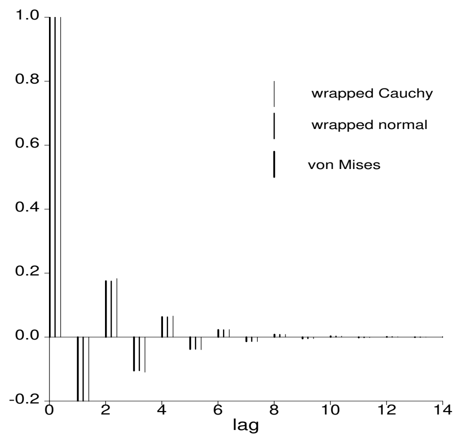

Figure 14: Cacf in several circular HMMs, parameter values $(\mu_1, \mu_2) = (1, 3)$, $\gamma_{1,2} = \gamma_{2,1} = 0.8)$, $\kappa_1 = \kappa_2 = 4$ (von Mises HMM) and $\rho_1 = \rho_2 = 0.86$ (wrapped Cauchy and normal HMM)

Hajo Holzmann
Institut für mathematische Stochastik
Georg-August-Universität Göttingen
Lotzestr. 13
37083 Göttingen
Germany
email: holzmann@math.uni-goettingen.de

Adress for correspondence:

Prof. Dr. Axel Munk
Institut für mathematische Stochastik
Georg-August-Universität Göttingen
Lotzestr. 13
37083 Göttingen
Germany
email: munk@math.uni-goettingen.de

Dr. Max Suster
McGill Centre for Research in Neuroscience
Montreal General Hospital
Montreal H3G 1A4
Canada
email: msuster@utm.utoronto.ca

Prof. Dr. Walter Zucchini
Institut für Statistik und Ökonometrie
Georg-August-Universität Göttingen
Platz der Göttinger Sieben 5
37073 Göttingen
Germany
email: wzucchi@uni-goettingen.de

Hajo Holzmann is PhD student and research associate at the Georg-August-University Göttingen, where he also got his Diploma in Mathematics and Economics in 2001. His research interest focuses on subjects in probability and statistics, with special emphasis on distributional limit theorems for dynamical systems, hidden Markov and mixture models and circular data analysis. This author acknowledges financial support of the Graduiertenkolleg "Gruppen und Geometrie".

Axel Munk is Professor of Mathematics and head of the Institute of Mathematical Stochastics, Georg-August-University Göttingen. He received his PhD at Göttingen University in 1994. From 1994 to 1999 he was Assistant Professor at the Ruhr-University Bochum and from 2000-2002 Associate Professor at the University of Paderborn. His research interest is Statistics and its Applications, in particular statistical modeling, medical statistics and curve estimation. A. Munk is currently involved in a project on estimating diffusion coefficients in groundwater problems.

Max Suster is a postdoctoral fellow at the McGill Centre for research in neuroscience, Montreal, Canada. He received his PhD from the University of Cambridge, England, in 2001 and pursued his postdoctoral research at the University of Toronto, Canada, during 2001-2002. His research interests focus on the neural control of locomotion, with particular emphasis on mechanisms underlying the development of motor circuitry in genetically tractable organisms. He is currently investigating neural mechanisms of locomotor development in zebrafish.

Walter Zucchini is Professor of Statistics and head of the Institute of Statistics and Econometrics, Georg–August–University, Göttingen. He received his PhD in 1978 at the University of Natal, Durban, South Africa. He has held appointments at the Universities of Natal, Stellenbosch and Cape Town (South Africa), Roma (Lesotho). His research interests include model selection, time series analysis and statistical applications in hydrology, estimation of animal abundance, forestry and finance.