CHAPTER 12

Wind direction at Koeberg

12.1 Introduction

South Africa's only nuclear power station is situated at Koeberg on the west coast, about 30 km north of Cape Town. Wind direction, wind speed, rainfall and other meteorological data are collected continuously by the Koeberg weather station with a view to their use in radioactive plume modelling, *inter alia*. Four years of data were made available by the staff of the Koeberg weather station, and this chapter describes an attempt to model the wind direction at Koeberg by means of HMMs.

The wind direction data consist of hourly values of average wind direction over the preceding hour at 35 m above ground level. The period covered is 1 May 1985 to 30 April 1989 inclusive. The average referred to is a vector average, which allows for the circular nature of the data, and is given in degrees. There are in all 35 064 observations; there are no missing values.

12.2 Wind direction classified into 16 categories

Although the hourly averages of wind direction were available in degrees, the first group of models fitted treated the observations as lying in one of the 16 conventional directions N, NNE, ..., NNW, coded 1 to 16 in that order. This was done in order to illustrate the application of HMMs to time series of categorical observations, a special class of multinomial-like time series.

12.2.1 Three HMMs for hourly averages of wind direction

The first model fitted was a simple multinomial–HMM with two states and no seasonal components, the case $m = 2$ and $q = 16$ of the categorical model described in Section 8.4.2. In this model there are 32 parameters to be estimated: two transition probabilities to specify the Markov chain, and 15 probabilities for each of the two states, subject to the sum of the 15 not exceeding one. The results are as follows. The underlying Markov chain has transition probability matrix $\begin{pmatrix} 0.964 & 0.036 \\ 0.031 & 0.969 \end{pmatrix}$ and stationary distribution (0.462, 0.538), and the 16 probabilities associated with each

Table 12.1 *Koeberg wind data (hourly):* 1000×*probabilities of each direction in each state for (from left to right) the simple two-state HMM, the simple three-state HMM, and the two-state HMM with cyclical components.*

		2-state HMM		3-state HMM			cyclic HMM	
1	N	129	0	148	0	1	127	0
2	NNE	48	0	47	0	16	47	0
3	NE	59	1	16	0	97	57	2
4	ENE	44	26	3	0	148	27	40
5	E	6	50	1	0	132	4	52
6	ESE	1	75	0	0	182	1	76
7	SE	0	177	0	23	388	1	179
8	SSE	0	313	0	426	33	0	317
9	S	1	181	0	257	2	1	183
10	SSW	4	122	2	176	0	7	121
11	SW	34	48	20	89	0	59	26
12	WSW	110	8	111	28	0	114	3
13	W	147	0	169	2	0	145	0
14	WNW	130	0	151	0	0	128	0
15	NW	137	0	159	0	1	135	0
16	NNW	149	0	173	0	0	147	0

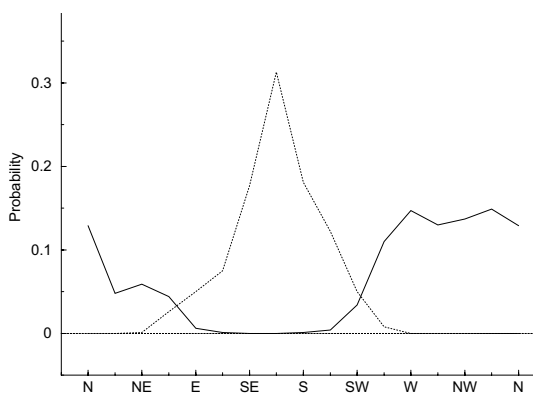

Figure 12.1 *Koeberg wind data (hourly): probabilities of each direction in the simple two-state HMM.*

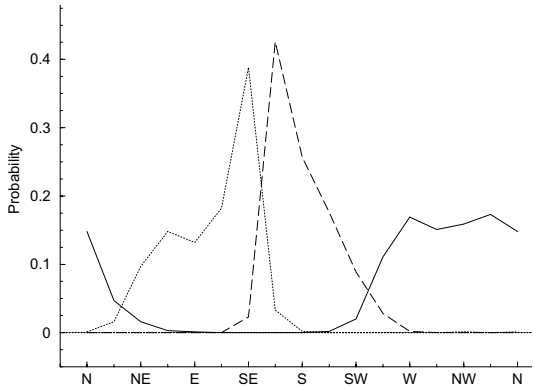

Figure 12.2 *Koeberg wind data (hourly): probabilities of each direction in the three-state HMM.*

of the two states are displayed in columns 3 and 4 of Table 12.1. A graph of these two sets of probabilities appears as Figure 12.1. The value of the unconditional log-likelihood achieved by this model is −75 832.1.

The model successfully identifies two apparently meaningful weather states which are very different at least as regards the likely wind directions in those states. In state 1 the most likely direction is NNW, and the probability falls away on either side of NNW, reaching a level of less than 0.01 for all directions (clockwise) from E to SSW inclusive. In state 2 the peak is at direction SSE, and the probability falls away sharply on either side, being less than 0.01 for all directions from WSW to NE inclusive. Two generalizations of this type of model were also fitted: firstly, a model based on a three-state Markov chain rather than a two-state one; and secondly, a model based on a two-state Markov chain but incorporating both a daily cycle and an annual cycle. We shall now describe these two models, and in due course compare all the models considered on the basis of AIC and BIC.

The three-state HMM has 51 parameters: six to specify the Markov chain, and 15 probabilities associated with each of the states. Essentially this model splits state 2 of the two-state model into two new states, one of which peaks at SSE and the other at SE. The transition probability matrix is

$$\begin{pmatrix} 0.957 & 0.030 & 0.013 \\ 0.015 & 0.923 & 0.062 \\ 0.051 & 0.077 & 0.872 \end{pmatrix}$$

and the stationary distribution is (0.400, 0.377, 0.223). The 16 probabilities associated with each of the three states are displayed in Fig-

ure 12.2 and columns 5–7 of Table 12.1. The unconditional log-likelihood is −69 525.9.

As regards the model which adds daily and annual cyclical effects to the simple two-state HMM, it was decided to build these effects into the Markov chain rather than into the state-dependent probabilities. This was because, for a two-state chain, we can model cyclical effects parsimoniously by assuming that the two off-diagonal transition probabilities

$$\Pr(C_t \neq i \mid C_{t-1} = i) = {}_t\gamma_{i,3-i} \quad (i = 1, 2)$$

are given by an appropriate periodic function of t. We assume that $\text{logit}\,{}_t\gamma_{i,3-i}$ is equal to

$$a_i + \underbrace{b_i \cos(2\pi t/24) + c_i \sin(2\pi t/24)}_{\text{daily cycle}} + \underbrace{d_i \cos(2\pi t/8766) + e_i \sin(2\pi t/8766)}_{\text{annual cycle}} \tag{12.1}$$

for $i = 1, 2$ and $t = 2, 3, \ldots, T$. A similar model for each of the state-dependent probabilities in each of the two states would involve many more parameters. As discussed in Section 8.5.2, the estimation technique has to be modified when the underlying Markov chain is not assumed to be homogeneous. The estimates in this case were based on the initial state of the Markov chain being state 2, and the log of the likelihood conditioned on that initial state is −75 658.5. (Conditioning on state 1 yielded a slightly inferior value for the likelihood, and similar parameter estimates.)

The probabilities associated with each state are given in the last two columns of Table 12.1, and the estimated parameters for the two off-diagonal transition probabilities are given below, in the notation of Equation (12.1).

i	a_i	b_i	c_i	d_i	e_i
1	−3.349	0.197	−0.695	−0.208	−0.401
2	−3.523	−0.272	0.801	0.082	−0.089

From Table 12.1 it will be noted that the general pattern of the state-dependent probabilities is in this case very similar to that of the simple two-state model without any cyclical components.

12.2.2 Model comparisons and other possible models

The three models described above were compared with each other and with a saturated 16-state Markov chain model, on the basis of AIC and BIC. The transition probabilities defining the Markov chain model were estimated by conditional maximum likelihood, as described in Section

Table 12.2 *Koeberg wind data (hourly):* 1000×*transition probability matrix of saturated Markov chain model.*

610	80	22	8	3	1	1	2	4	0	3	3	14	20	37	190
241	346	163	37	15	1	3	3	3	3	9	10	29	37	36	64
56	164	468	134	28	6	9	6	4	6	6	18	32	18	19	25
13	33	163	493	144	32	17	11	11	17	7	12	14	14	10	10
9	7	48	249	363	138	53	27	33	22	11	6	8	7	13	9
4	9	10	51	191	423	178	51	25	23	8	7	4	6	4	4
1	1	3	6	23	141	607	160	36	8	6	3	2	1	1	1
1	1	1	3	5	16	140	717	94	13	5	3	2	1	0	0
4	1	2	4	5	8	25	257	579	77	17	9	5	3	1	0
2	2	2	1	6	5	10	36	239	548	93	41	10	5	2	0
5	2	3	5	3	3	8	12	38	309	397	151	38	12	8	5
4	2	2	3	1	3	2	5	17	56	211	504	149	19	16	7
10	5	5	3	4	1	2	5	4	13	28	178	561	138	30	13
13	5	4	3	3	3	1	1	1	7	8	27	188	494	199	43
31	9	7	4	5	2	1	1	2	1	3	11	43	181	509	190
158	23	9	5	1	1	2	1	0	2	2	4	17	54	162	559

Table 12.3 *Koeberg wind data (hourly): comparison of four models fitted.*

model	k	$-l$	AIC	BIC
2-state HMM	32	75 832.1	151 728	151 999
3-state HMM	51	69 525.9	139 154	139 585
2-state HMM with cycles	40	75 658.5*	151 397	151 736
saturated Markov chain	240	48 301.7	**97 083**	**99 115**

* conditional on state 2 being the initial state

1.3.5, and are displayed in Table 12.2. The comparison appears as Table 12.3.

What is of course striking in Table 12.3 is that the likelihood of the saturated Markov chain model is so much higher than that of the HMMs that the large number of parameters of the Markov chain (240) is virtually irrelevant when comparisons are made by AIC or BIC. It is therefore interesting to compare certain properties of the Markov chain model with the corresponding properties of the simple two-state HMM (e.g. unconditional and conditional probabilities). The unconditional probabilities of each direction were computed for the two models, and are almost identical.

However, examination of $\Pr(X_{t+1} = 16 \mid X_t = 8)$, where X_t denotes the direction at time t, points to an important difference between the

Table 12.4 *Koeberg wind data (daily at hour 1): comparison of two models fitted.*

model	k	$-l$	AIC	BIC
2-state HM	32	3461.88	**6987.75**	**7156.93**
saturated MC	240	3292.52	7065.04	8333.89

models. For the Markov chain model this probability is zero, since no such transitions were observed. For the HMM it is, in the notation of Equation (8.1),

$$\frac{\boldsymbol{\delta}\mathbf{\Pi}(8)\mathbf{\Gamma}\mathbf{\Pi}(16)\mathbf{1}'}{\boldsymbol{\delta}\mathbf{\Pi}(8)\mathbf{1}'} = \frac{0.5375 \times 0.3131 \times 0.0310 \times 0.1494}{0.5375 \times 0.3131} = 0.0046.$$

Although small, this probability is not insignificant. The observed number of transitions from SSE (direction 8) was 5899. On the basis of the HMM one would therefore expect about 27 of these transitions to be to NNW. None were observed. Under the HMM 180-degree switches in direction are quite possible. Every time the state changes (which happens at any given time point with probability in excess of 0.03), the most likely direction of wind changes by 180 degrees. This is inconsistent with the observed gradual changes in direction; the matrix of observed transition counts is heavily dominated by diagonal and near-diagonal elements (and, because of the circular nature of the categories, elements in the corners furthest from the principal diagonal). Changes through 180 degrees were in general rarely observed.

The above discussion suggests that, if daily figures are examined rather than hourly, the observed process will be more amenable to modelling by means of an HMM, because abrupt changes of direction are more likely in daily data than hourly. A Markov chain and a two-state HMM were therefore fitted to the series of length 1461 beginning with the first observation and including every 24th observation thereafter. For these data the HMM proved to be superior to the Markov chain even on the basis of AIC, which penalizes extra parameters less here than does BIC: see Table 12.4. Although a daily model is of little use in the main application intended (evacuation planning), there are other applications for which a daily model is exactly what one needs, e.g. forecasting the wind direction a day ahead.

Since the (first-order) Markov chain model does not allow for dependence beyond first order, the question that arises is whether any model for the hourly data which allows for higher-order dependence is superior to the Markov chain model; the HMMs considered clearly are not.

Table 12.5 *Koeberg wind data (hourly): comparison of first-order Markov chain with lag-2 Raftery models.*

model	k	$-l$	AIC	BIC
saturated MC	240	48 301.7	97 083.4	99 115.0
Raftery model with starting values	241	48 087.8*	96 657.5	98 697.6
Raftery model fitted by max. likelihood	241	48 049.8*	**96 581.6**	**98 621.7**

* conditioned on the first two states

A saturated second-order Markov chain model would have an excessive number of parameters, and the Pegram model for (e.g.) a second-order Markov chain cannot reflect the property that, if a transition is made out of a given category, a nearby category is a more likely destination than a distant one.

The Raftery models (also known as MTD models: see p. 23) do not suffer from that disadvantage, and in fact it is very easy to find a lag-2 Raftery model that is convincingly superior to the first-order Markov chain model. In the notation of Section 1.3.6, take $\mathbf{Q}$ to be the transition probability matrix of the first-order Markov chain, as displayed in Table 12.2, and perform a simple line-search to find that value of λ_1 which maximizes the resulting conditional likelihood. This turns out to be 0.925. With these values for $\mathbf{Q}$ and λ_1 as starting values, the conditional likelihood was then maximized with respect to all 241 parameters, subject to the assumption that $0 \leq \lambda_1 \leq 1$. (This assumption makes it unnecessary to impose 16^3 pairs of nonlinear constraints on the maximization, and seems reasonable in the context of hourly wind directions showing a high degree of persistence.) The resulting value for λ_1 is 0.9125, and the resulting matrix $\mathbf{Q}$ does not differ much from its starting value. The log-likelihood achieved is $-48\,049.8$: see Table 12.5 for a comparison of likelihood values and the usual model selection criteria, from which the Raftery model emerges as superior to the Markov chain. Berchtold (2001) reports that this model cannot be improved on by another MTD model.

12.2.3 Conclusion

The conclusion we may draw from the above analysis is that, both for hourly and daily (categorized) wind direction data, it is possible to fit a model which is superior (according to the model selection criteria used)

to a saturated first-order Markov chain. In the case of the hourly data, a lag-2 Raftery model (i.e. a particular kind of second-order Markov chain model) is preferred. In the case of the daily data, a simple two-state HMM for categorical time series, as introduced in Section 8.4.2, performs better than the Markov chain.

12.3 Wind direction as a circular variable

We now revisit the Koeberg wind direction data, with a view to demonstrating some of the many variations of HMMs that could be applied to data such as these. In particular, we do not confine ourselves here to use of the directions as classified into the 16 points of the compass, and we make use of the corresponding observations of wind speed in several different ways as a covariate in models for the direction, or change in direction. Some of the models are not particularly successful, but it is not our intention here to confine our attention to models which are in some sense the best, but rather to demonstrate the versatility of HMMs and to explore some of the questions which may arise in the fitting of more complex models.

12.3.1 Daily at hour 24: von Mises–HMMs

The analysis of wind direction that we have so far described is based entirely on the data as classified into the 16 directions N, NNE, ..., NNW. Although it illustrates the fitting of models to categorical series, it does ignore the fact that the observations of (hourly average) direction are available in degrees, and it ignores the circular nature of the observations.

One of the strengths of the hidden Markov formulation is that almost any kind of data can be accommodated at observation level. Here we can exploit that flexibility by taking the state-dependent distribution from a family of distributions designed for circular data, the von Mises distributions. The probability density function of the von Mises distribution with parameters μ and κ is

$$f(x) = (2\pi I_0(\kappa))^{-1} \exp(\kappa \cos(x - \mu)) \quad \text{for } x \in (-\pi, \pi];$$

see also Section 11.2.

In this section we shall fit to the directions data (in degrees) HMMs which have 1–4 states and von Mises state-dependent distributions. This is perhaps a fairly obvious approach if one has available 'continuous' circular data, e.g. in integer degrees. We do this for the series of length 1461 formed by taking every 24th observation of direction. By taking

Table 12.6 *Koeberg wind data (daily at hour 24): comparison of models fitted to wind direction. m is the number of states in an HMM, k is in general the number of parameters in a model, and equals $m^2 + m$ for all but the Markov chain model. Note that the figures in the $-l$ column for the continuous models have to be adjusted before they can be compared with those for the discretized models.*

model	m	k	$-l$	$-l$(adjusted)	AIC	BIC
von Mises–HM	1	2	2581.312	3946.926	7897.9	7908.4
von Mises–HM	2	6	2143.914	3509.528	7031.1	7062.8
von Mises–HM	3	12	2087.133	3452.747	6929.5	6992.9
von Mises–HM	4	20	2034.133	3399.747	**6839.5**	**6945.2**
discretized von M–HM	1	2	3947.152		7898.3	7908.9
discretized von M–HM	2	6	3522.848		7057.7	7089.4
discretized von M–HM	3	12	3464.491		6953.0	7016.4
discretized von M–HM	4	20	3425.703		6891.4	6997.1
saturated Markov chain		240	3236.5		6953.0	8221.9

every 24th observation we are focusing on the hour each day from 23:00 to 24:00, and modelling the average direction over that hour.

It is somewhat less obvious that it is still possible to fit von Mises–HMMs if one has available only the categorized directions; that is a second family of models which we present here. To fit such models, one uses in the likelihood computation not the von Mises density but the integral thereof over the interval corresponding to the category observed. For the purpose of comparison we present also a 16-state saturated Markov chain fitted to the same series.

A minor complication that arises if one wishes to compare the continuous log-likelihoods with the discrete is that one has to bear in mind that the direct use of a density in a continuous likelihood is just a convention. The density should in fact be integrated over the smallest interval that can contain the observation, or the integral approximated by the density multiplied by the appropriate interval length. The effect here is that, if one wishes to compare the continuous log-likelihoods with those relating to to the 16 categories, one has to add $1461 \log(2\pi/16) = -1365.614$ to the continuous log-likelihood, or equivalently subtract it from minus the log-likelihood l. The resulting adjusted values of $-l$ appear in the column headed '$-l$(adjusted)' in Table 12.6.

12.3.2 Modelling hourly change of direction

Given the strong persistence of wind direction which is apparent from the transition probability matrix in Table 12.2, the most promising approach, however, seems to be to model the *change* in direction rather then the direction itself. We therefore describe here a variety of models for the change in (hourly average) direction from one hour to the next, both with and without the use of wind speed, lagged one hour, as a covariate. Observations of wind speed, in cm s^{-1}, were available for the same period as the observations of direction. The use of lagged wind speed as a covariate, rather than simultaneous, is motivated by the need for any covariate to be available at the time of forecast; here a 1-hour-ahead forecast is of interest.

12.3.3 Transition probabilities varying with lagged speed

Change of direction, like direction itself, is a circular variable, and first we model it by means of a von Mises–HMM with two, three, or four states. Then we introduce wind speed, lagged one hour, into the model as follows.

The usual transformation of the transition probabilities in row i (say) is

$$\gamma_{ij} = e^{\tau_{ij}} \Big/ \left(1 + \textstyle\sum_{k \neq i} e^{\tau_{ik}}\right) \quad j \neq i,$$

with γ_{ii} then determined by the row-sum constraint. Here we use instead the transformation

$$\gamma_{ij} = \Pr(C_t = j \mid C_{t-1} = i, S_{t-1} = s) = e^{\tau_{ij}} \Big/ \textstyle\sum_{k=1}^{m} e^{\tau_{ik}} \quad j = 1, 2, \ldots m$$

where

$$\tau_{ii} = \eta_i s$$

and S_{t-1} is the speed at time $t-1$. Equivalently, the transformation is, for row i,

$$\gamma_{ij} = e^{\tau_{ij}} \Big/ \left(e^{\eta_i s} + \textstyle\sum_{k \neq i} e^{\tau_{ik}}\right) \quad j \neq i,$$

with γ_{ii} determined by the row-sum constraint. This structure allows the speed at time $t-1$ to influence the probabilities of transition from state i to j between times $t-1$ and t.

In passing, we note that there would be no point in introducing an intercept η_{0i} as follows:

$$\tau_{ii} = \eta_{0i} + \eta_i s.$$

In all the corresponding transition probabilities, η_{0i} would be confounded with the parameters τ_{ij} and therefore non-identifiable. This can be seen as follows. With $\tau_{ii} = \eta_{0i} + \eta_i s$, γ_{ii} would then be given by

$$\gamma_{ii} = e^{\eta_i s} \Big/ \left(e^{\eta_i s} + \textstyle\sum_{k \neq i} e^{\tau_{ik} - \eta_{0i}}\right),$$

Table 12.7 *Koeberg wind data (hourly): comparison of von Mises–HM models fitted to* change *in wind direction. The covariate, if any, is here used to influence the transition probabilities. The number of states is m, and the number of parameters k.*

covariate	m	k	$-l$	AIC	BIC
–	1	2	21821.350	43646.7	43663.6
–	2	6	8608.516	17229.0	17279.8
–	3	12	6983.205	13990.4	14092.0
–	4	20	6766.035	13572.1	13741.4
speed	2	8	6868.472	13752.9	13820.7
speed	3	15	5699.676	11429.4	11556.3
speed	4	24	5476.199	11000.4	11203.6
$\sqrt{\text{speed}}$	2	8	6771.533	13559.1	13626.8
$\sqrt{\text{speed}}$	3	15	5595.228	11220.5	11347.4
$\sqrt{\text{speed}}$	4	24	5361.759	**10771.5**	**10974.7**

and γ_{ij} (for $j \neq i$) by

$$\gamma_{ij} = e^{\tau_{ij}-\eta_{0i}} \Big/ \left(e^{\eta_i s} + \textstyle\sum_{k\neq i} e^{\tau_{ik}-\eta_{0i}}\right).$$

The results are displayed in Table 12.7, along with results for the corresponding models which use the square root of speed, rather than speed itself, as covariate. In these models, the square root of speed is in general more successful as a covariate than is speed.

12.3.4 Concentration parameter varying with lagged speed

A quite different way in which the square root of speed, lagged one hour, could be used as a covariate is via the concentration parameter (κ) of the von Mises distributions used here as the state-dependent distributions. The intuition underlying this proposal is that the higher the speed, the more concentrated will be the change in wind direction in the following hour (unless the state changes). One possibility is this: we assume that, given $S_{t-1} = s$, the concentration parameter in state i is

$$\log \kappa_i = \zeta_{i0} + \zeta_{i1}\sqrt{s}. \tag{12.2}$$

However, a model which seems both more successful (as judged by likelihood) and more stable numerically is this. Given $S_{t-1} = s$, let the concentration parameter in state i (not its logarithm) be a linear function of the square of the speed:

$$\kappa_i = \zeta_{i0} + \zeta_{i1}s^2. \tag{12.3}$$

Table 12.8 *Koeberg wind data (hourly): comparison of von Mises–HMMs fitted to change in wind direction. Here the covariate is used to influence the concentration parameter κ, as in Equation (12.3). The number of states is m, and the number of parameters k.*

covariate	m	k	$-l$	AIC	BIC
speed^2	1	3	10256.07	20518.1	20543.5
speed^2	2	8	4761.549	9539.1	9606.8
speed^2	3	15	4125.851	8281.7	8408.7
speed^2	4	24	3975.223	**7998.4**	**8201.6**

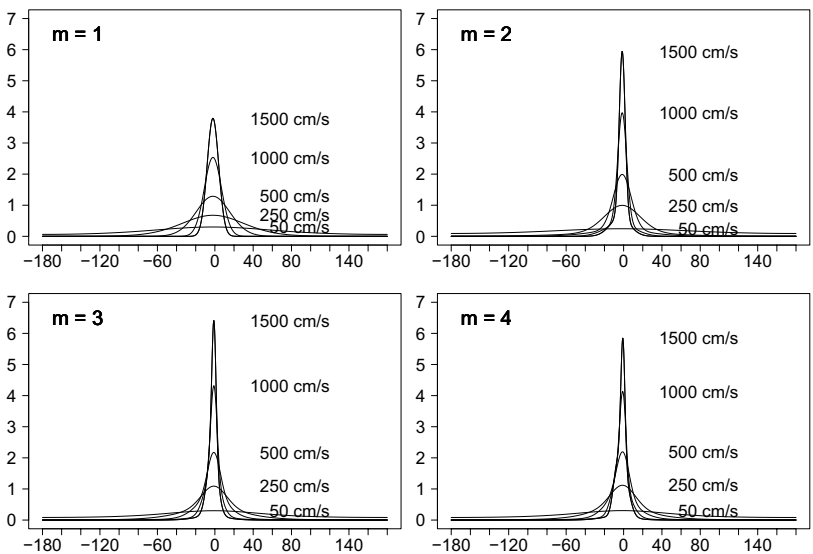

Figure 12.3 *Koeberg wind data (hourly), von Mises–HMMs of form (12.3): marginal distribution, for one to four states and several values of lagged speed, for change of direction (in degrees).*

To ensure that κ_i is positive we constrain ζ_{i0} and ζ_{i1} to be positive. Table 12.8 presents the log-likelihood values, AIC and BIC for four such models, and Figures 12.3 and 12.4 and Table 12.9 present some details of these models.

We have in this section considered a multiplicity of models for wind direction, but more variations yet would be possible. One could for in-

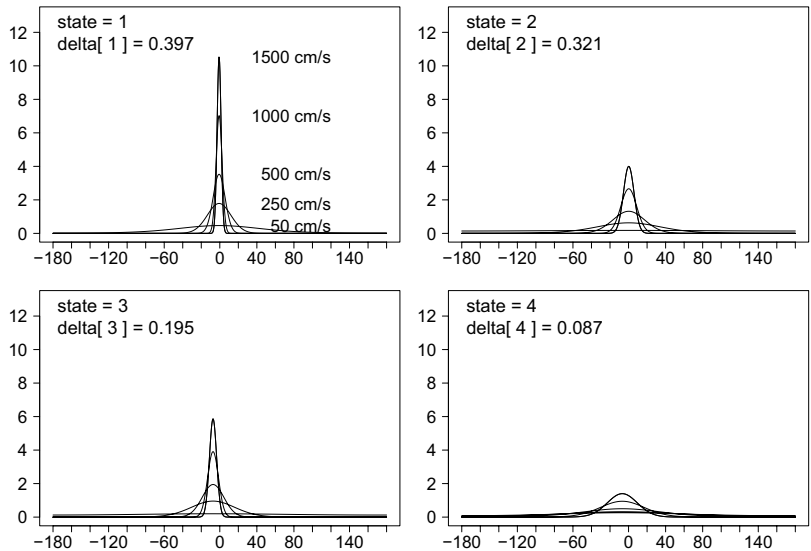

Figure 12.4 *Koeberg wind data (hourly), four-state von Mises–HMM of form (12.3): state-dependent distributions for change of direction (before multiplication by mixing probabilities δ_i).*

Table 12.9 *Koeberg wind data (hourly): parameters of four-state von Mises–HMM fitted to change in wind direction. Here the covariate is used to influence the concentration parameter κ, as in Equation (12.3).*

$$\mathbf{\Gamma} = \begin{pmatrix} 0.755 & 0.163 & 0.080 & 0.003 \\ 0.182 & 0.707 & 0.045 & 0.006 \\ 0.185 & 0.000 & 0.722 & 0.093 \\ 0.031 & 0.341 & 0.095 & 0.533 \end{pmatrix}$$

i	1	2	3	4
δ_i	0.397	0.321	0.195	0.087
μ_i	-0.0132	0.0037	-0.1273	-0.1179
ζ_{i0}	0.917	0.000	0.000	0.564
ζ_{i1}	31.01×10^{-5}	4.48×10^{-5}	9.61×10^{-5}	0.53×10^{-5}

stance allow the state-dependent location parameters μ_i (not the concentrations) of the change in direction to depend on the speed lagged one hour, or allow the concentrations to depend on the speed in different ways in the different states; there is no reason *a priori* why a single relation such as (12.3) should apply to all m states. In any application one would have to be guided very much by the intended use of the model.

Exercises

1.(a) Write an **R** function to generate a series of observations from an m-state categorical HMM with q categories, as described in Section 8.4.2. (Hint: modify the code in A.2.1.)

(b) Write functions to estimate the parameters of such a model. (Modify the code in A.1.)

(c) Generate a long series of observations using your code from 1(a), and estimate the parameters using your code from 1(b).

2. Use your code from 1(b) to fit two- and three-state models to the categorized wind directions data, and compare your models with those described in Section 12.2.1.

3. Generalize Exercise 1 to handle multinomial–HMMs, as described in Section 8.4.1, rather than merely categorical.