Hidden Markov models for circular-valued time series

Walter Zucchini

Institute for Statistics and Econmetrics
University of Göttingen

Joint work with: Hajo Holzmann & Axel Munk, University of Göttingen,
Iain MacDonald, University of Cape Town.

OUTLINE

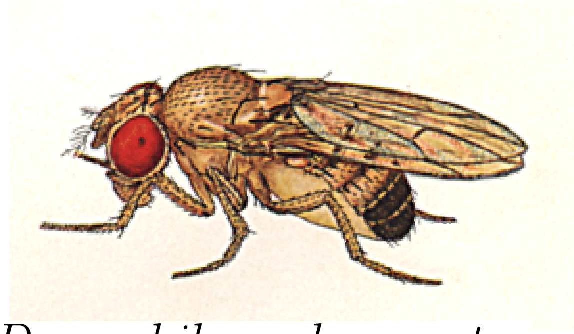
Part I – A simple HMM for circular-valued time series

1. Larval movements of *Drosophila*
2. Von Mises-HMMs
3. Some properties, methods to fit HMMs, and to assess the fit
4. Modelling speed and change of direction

Part II – Extensions of the simple HMM

1. Wind direction at Koeberg
2. A categorical-valued HMM
3. A discretized von Mises HMM
4. Modelling change of direction (cod)
5. Modelling cod using speed as a covariate

Larval movement of the fly Drosophila

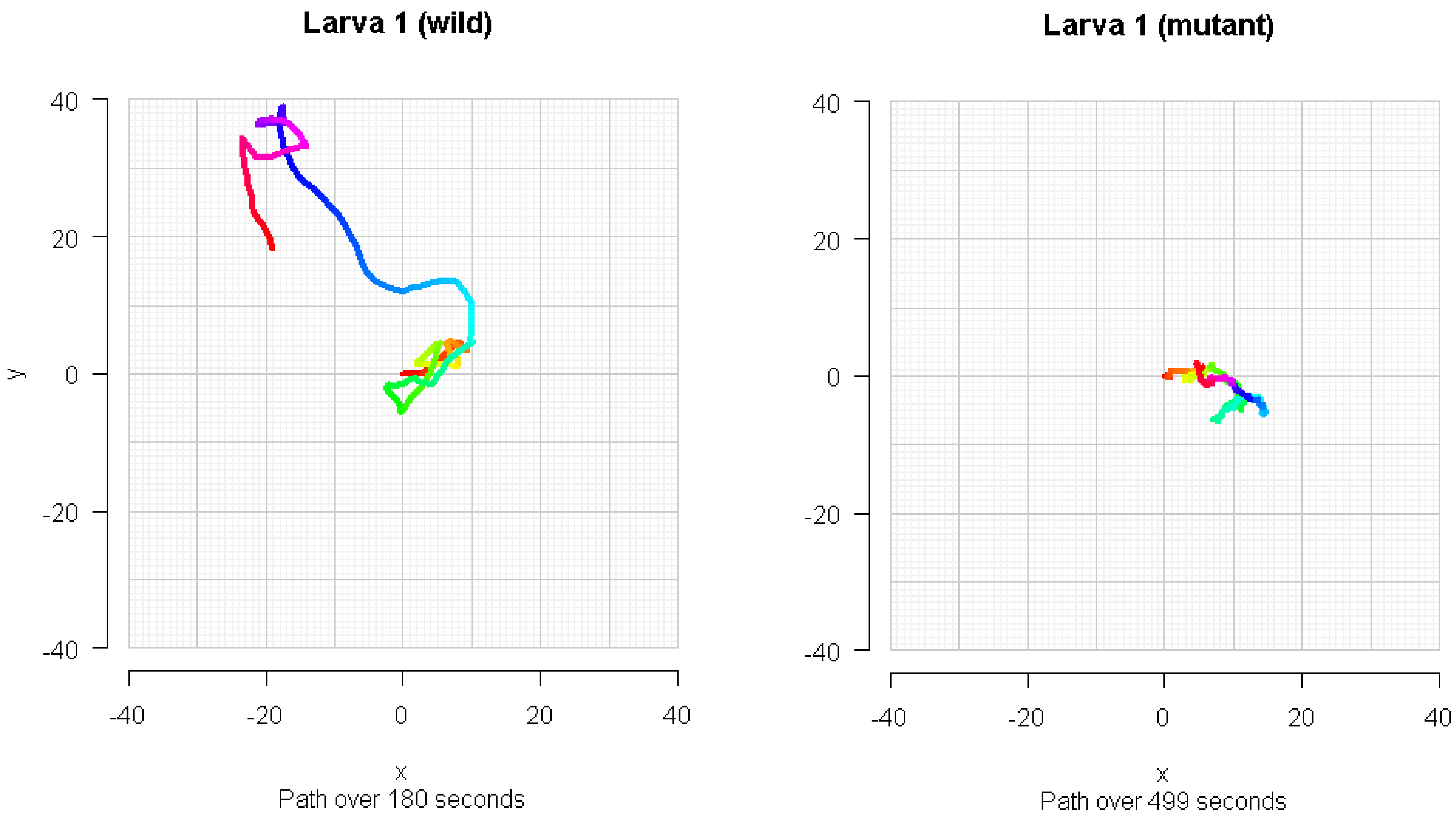


Drosophila melanogaster

Objective: Assess whether larvae that have been modified (mutants) behave differently from normal larvae (wild), and how.

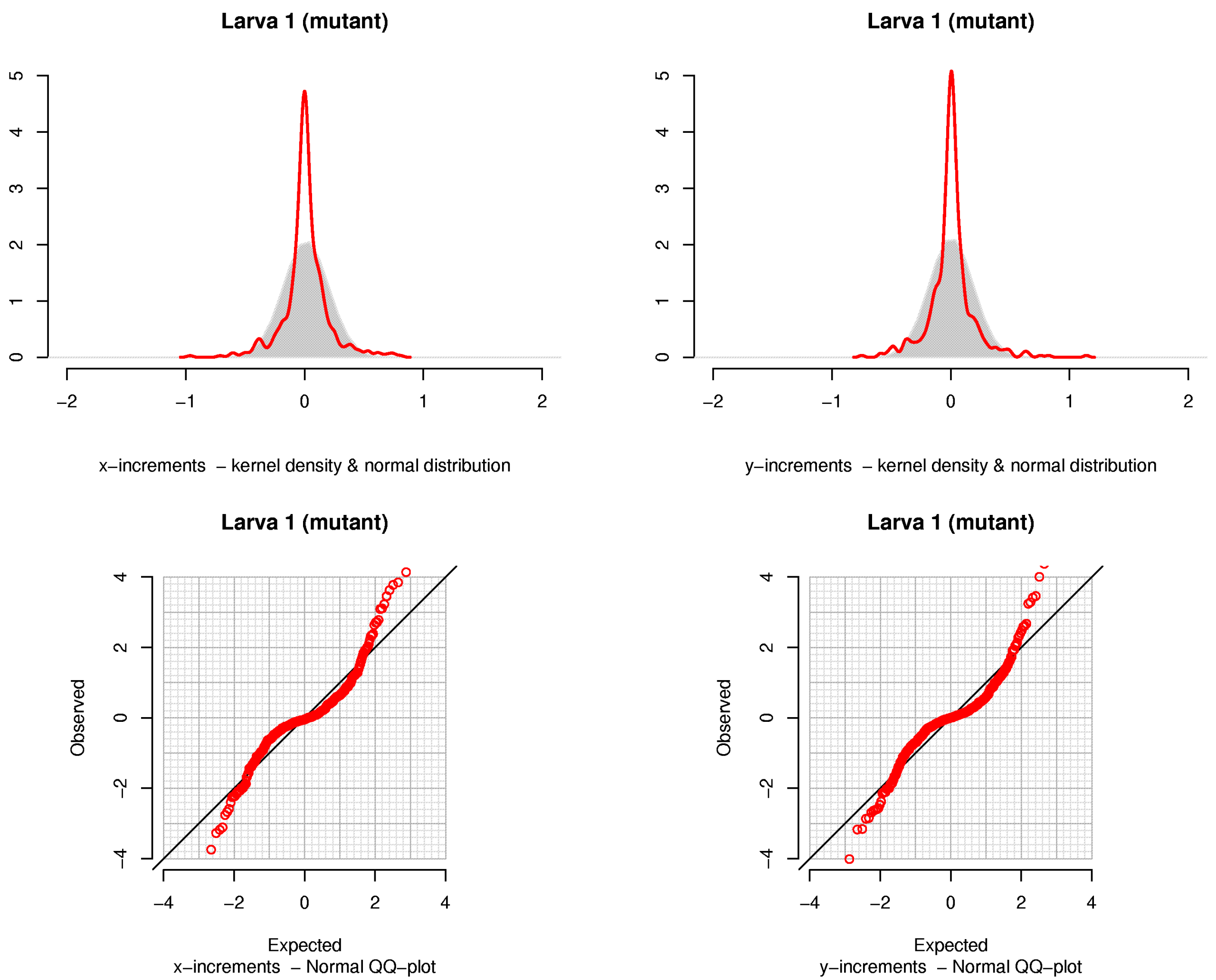
Data: Max Suster, McGill Centre for Research in Neuroscience
— 30 wild larvae with up to 180 observations each,
— 15 mutant larvae with up to 500 observations each.

Observations: Positions: $(x_1, y_1), (x_2, y_2), \ldots$ (resolution =1 sec)



The movements don't correspond to Brownian motion.

x-increments y-increments



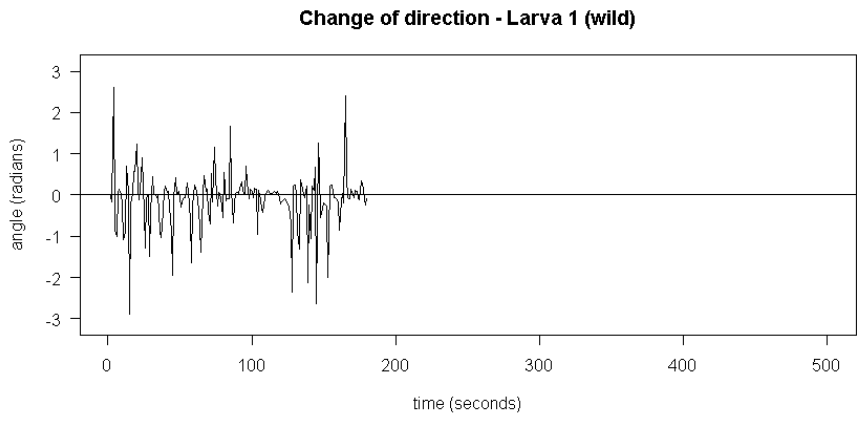
The increments are not the appropriate variables to model.

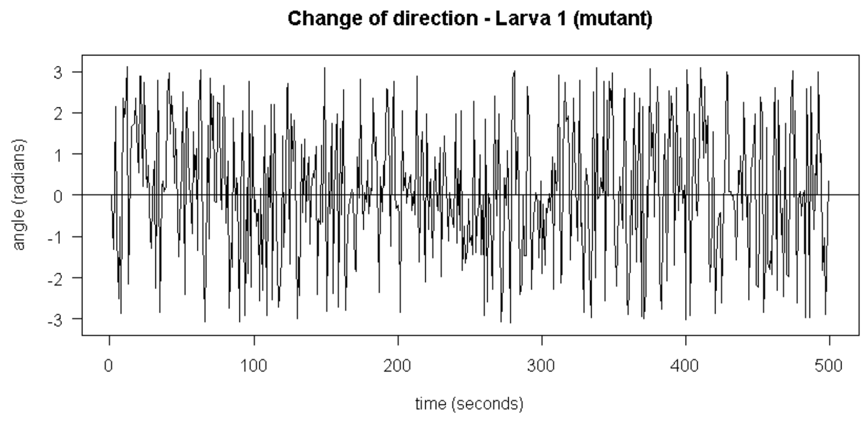
More promising is the bivariate time series (speed, change of direction).

variable	units	type
speed	mm per second	linear continuous–valued
change of direction	radians	circular–valued

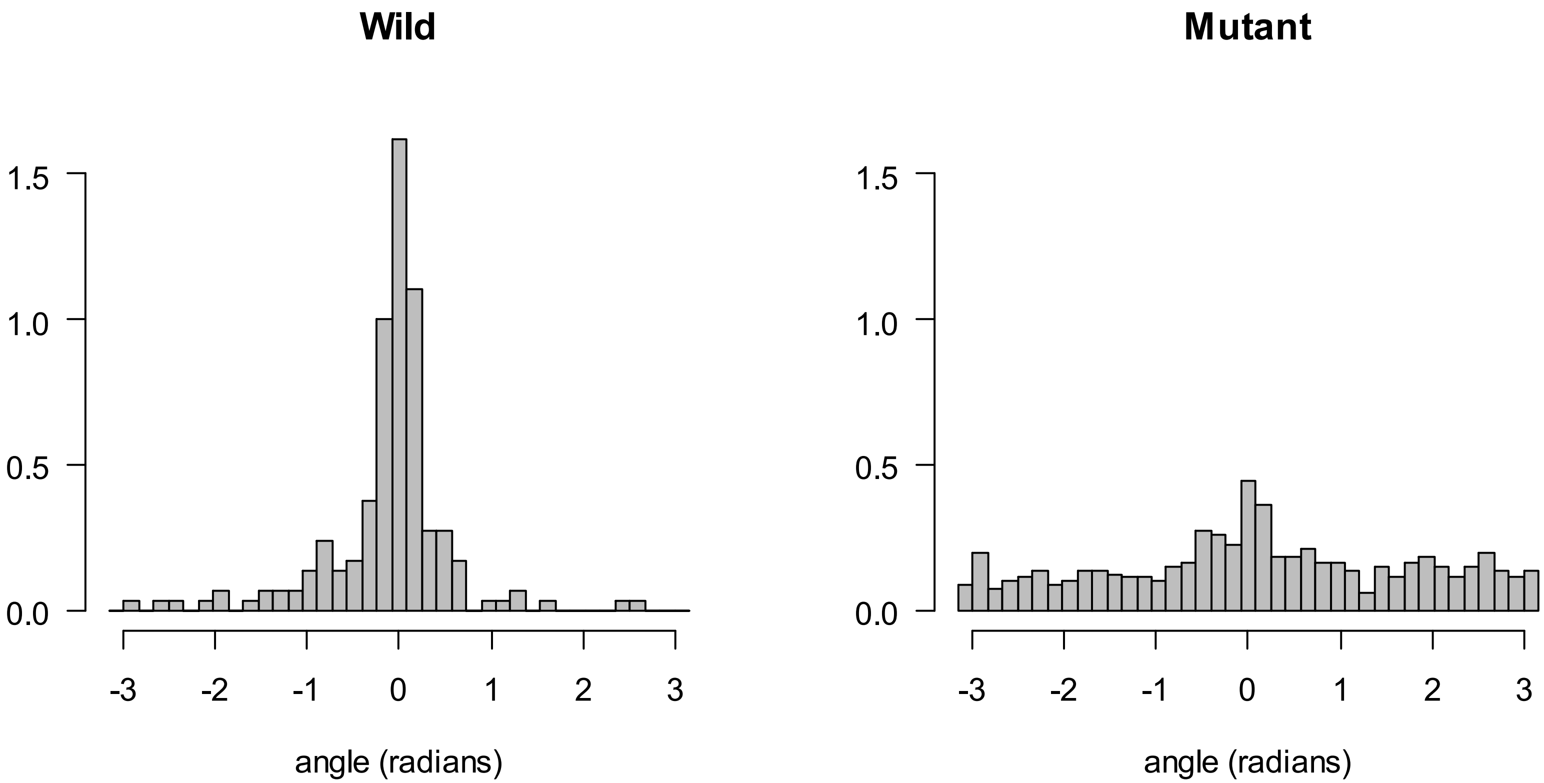
Modelling the time series change of direction (cod)

Time series: $a_1, a_2, \ldots, a_T$





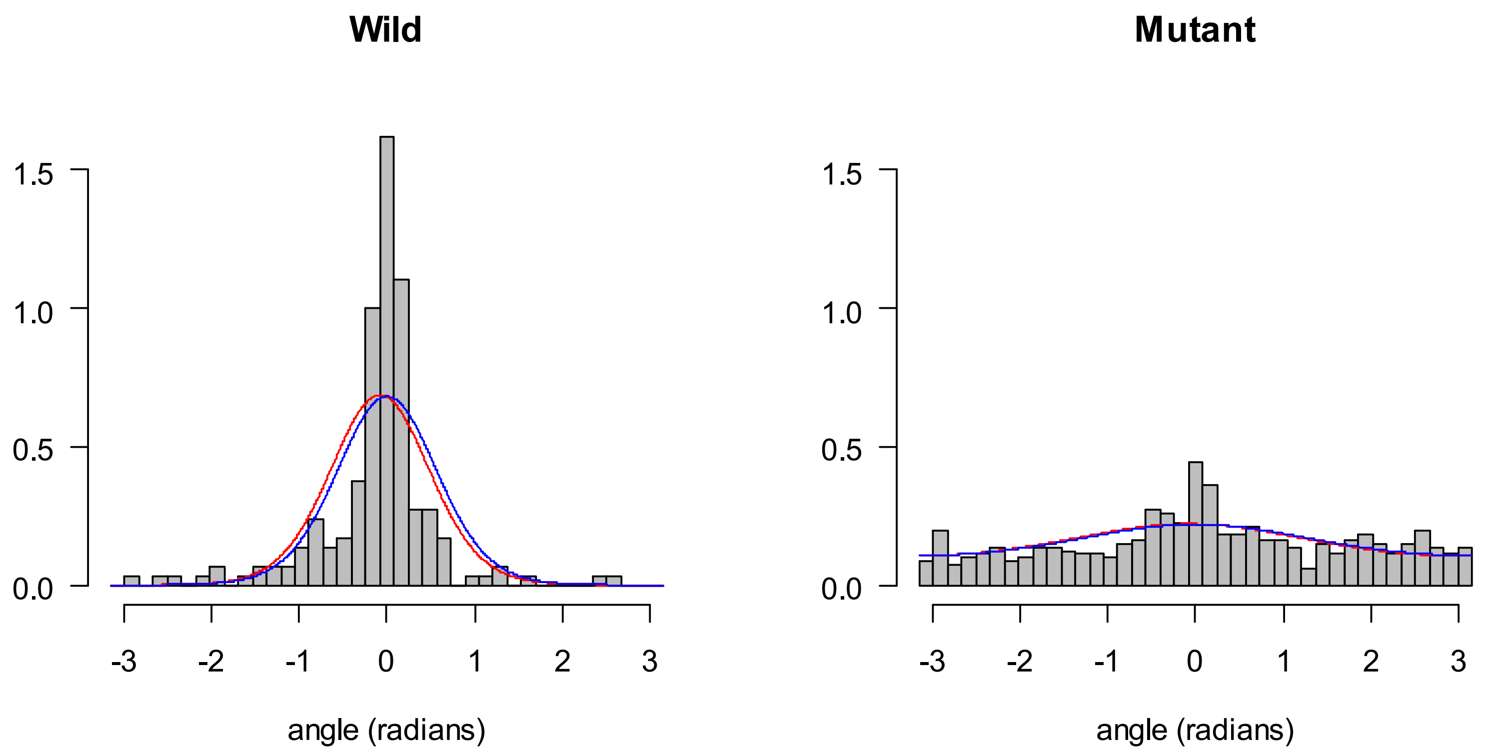
Histograms of change of direction



Drosophila

Von Mises distribution: $f(a) = \frac{1}{2\pi I_0(\kappa)} \; e^{\kappa \cos(a-\theta)} \qquad -\pi \le a < \pi$

Fitted von Mises distributions:



$\hat{\theta} = -0.08 \quad \hat{\kappa} = 3.27$

$\theta = 0 \quad \hat{\kappa} = 3.22$

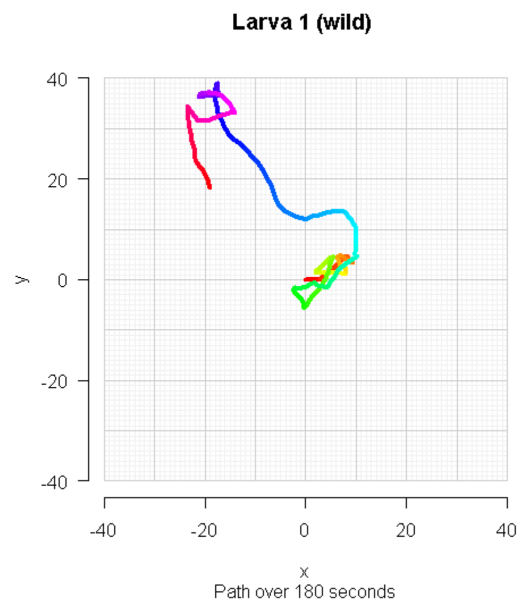
$\hat{\theta} = 0.22 \quad \hat{\kappa} = 0.37$

$\theta = 0 \quad \hat{\kappa} = 0.36$

The wrapped normal and wrapped Cauchy also fit poorly.

There are two types of movement:

type	speed	cod angle
turning (head-swinging)	low	large
linear	high	small



Mixture of two von Mises distributions:

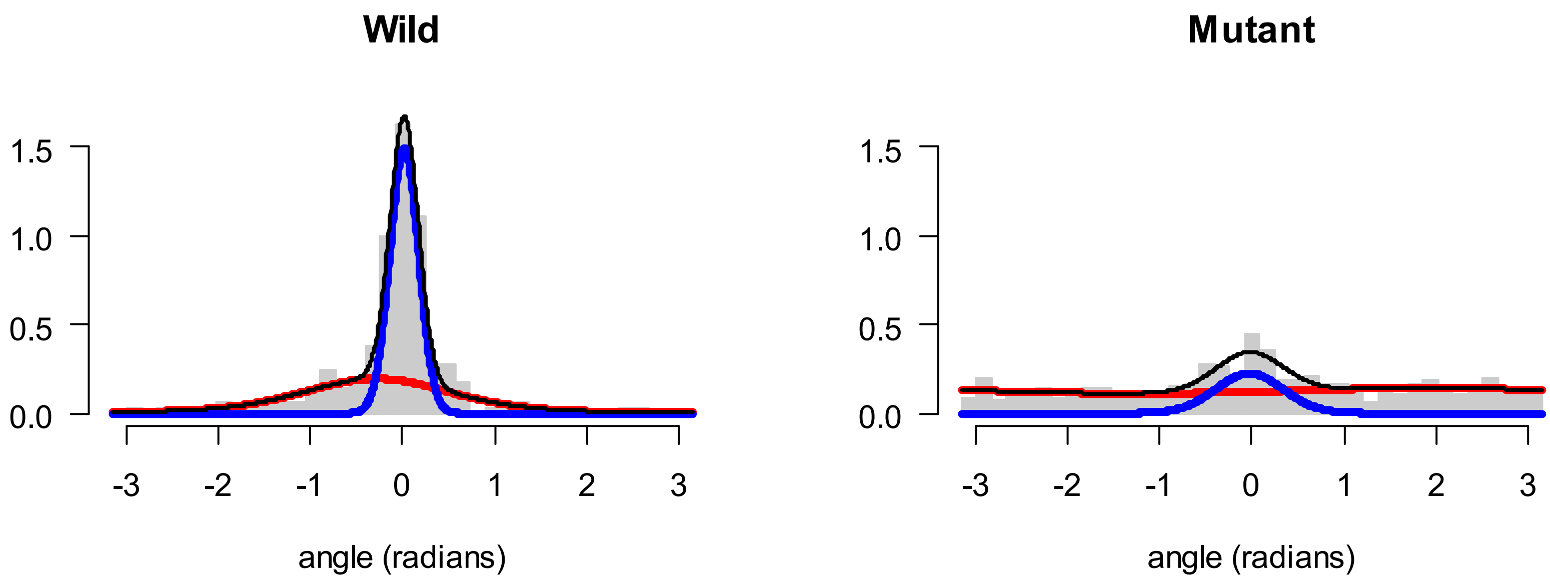
f_1 is $\mathrm{vM}(\theta_1, \kappa_1)$ with probability δ_1, (state 1)
f_2 is $\mathrm{vM}(\theta_2, \kappa_2)$ with probability δ_2, (state 2)

pdf of A: $f(a) = \delta_1 f_1(a) + \delta_2 f_2(a), \qquad \delta_1 + \delta_2 = 1$

The observations, $a_1, a_2, \ldots, a_T$, are generated in two stages:

parameter process: $C_1, C_2, \ldots, C_T$ determine the states,
state-dependent process: $A_1, A_2, \ldots, A_T$ observations, given the states.

Mixture of two von Mises distributions



f_1: vM(-0.28, 1.65) $\delta_1 = 0.42$
f_2: vM($0.02, 42.96$) $\delta_2 = 0.58$

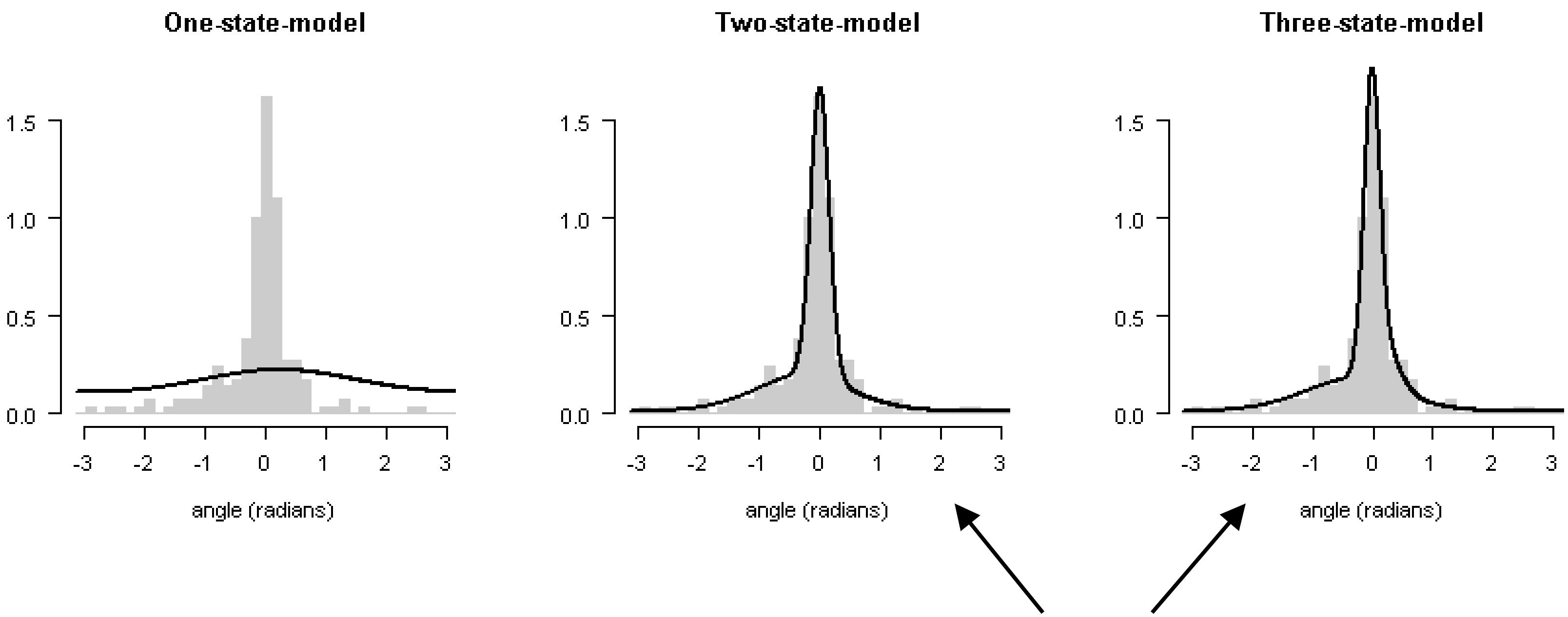
f_1: vM($1.91, 0.12$) $\delta_1 = 0.79$
f_2: vM($-0.03, 7.68$) $\delta_2 = 0.21$

The mixture (black) fits the marginal distribution much better.

Mixture of 3 von Mises distributions.

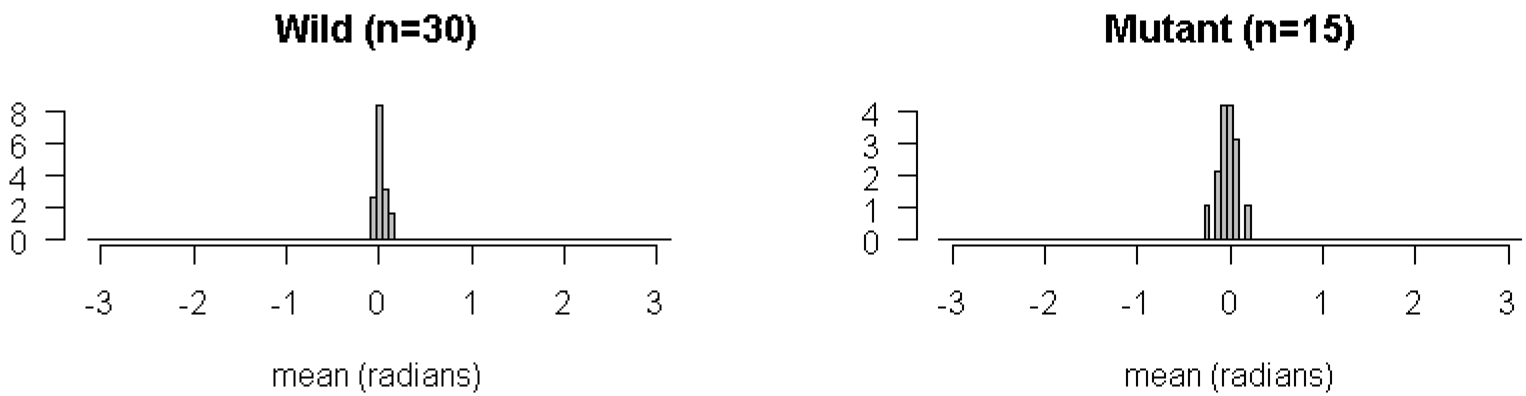
f_1 is $\mathrm{vM}(\theta_1, \kappa_1)$ with probability δ_1, (state 1)
f_2 is $\mathrm{vM}(\theta_2, \kappa_2)$ with probability δ_2, (state 2)
f_2 is $\mathrm{vM}(\theta_3, \kappa_3)$ with probability δ_3, (state 3)

$$f(a) = \delta_1 f_1(a) + \delta_2 f_2(a) + \delta_3 f_3(a), \qquad \delta_1 + \delta_2 + \delta_3 = 1$$



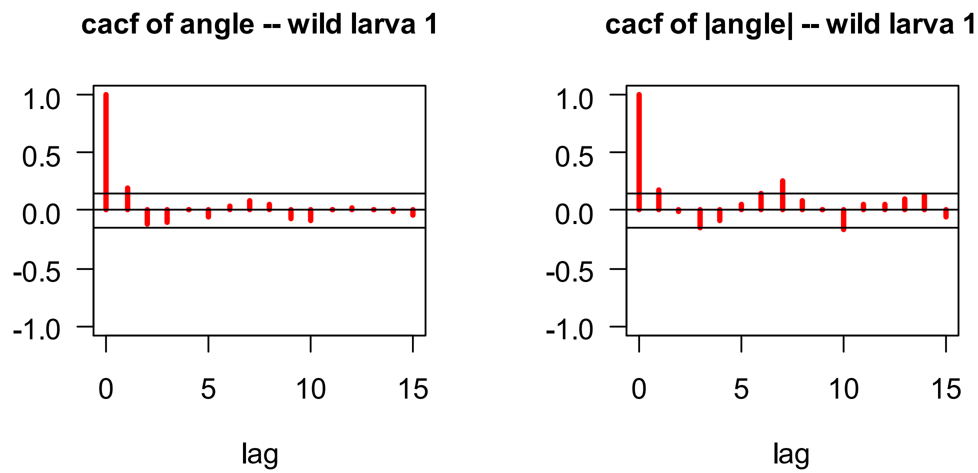
very similar fits

Histograms of circular means



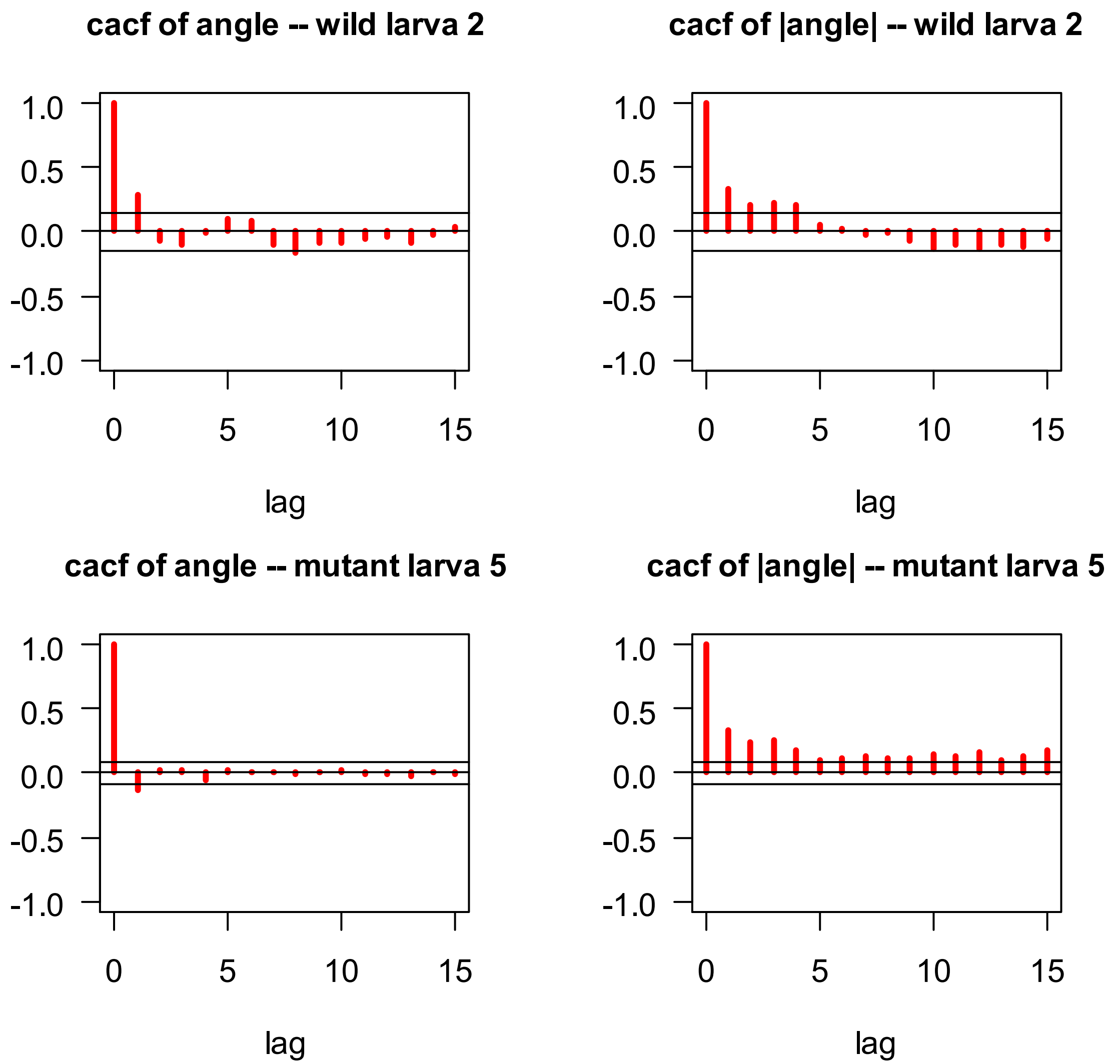
It makes little difference whether or not one sets $\theta = 0$.

The serial **circular correlation**[1] is small but there is serial **dependence**.



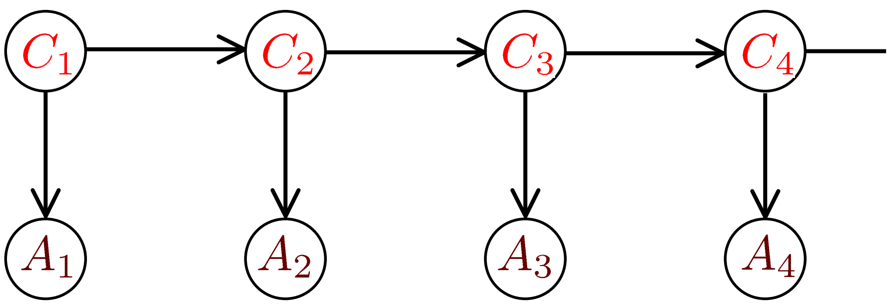
[1] Fisher and Lee (1994)

More examples



An HMM is a special kind of dependent mixture:

- Parameter process: $C_1, C_2, \cdots$ m-state Markov chain
- State-dependent process: $A_1, A_2, \cdots$ Observed process
- Assume conditional independence



Parameter process (hidden)

Observed state-dependent process

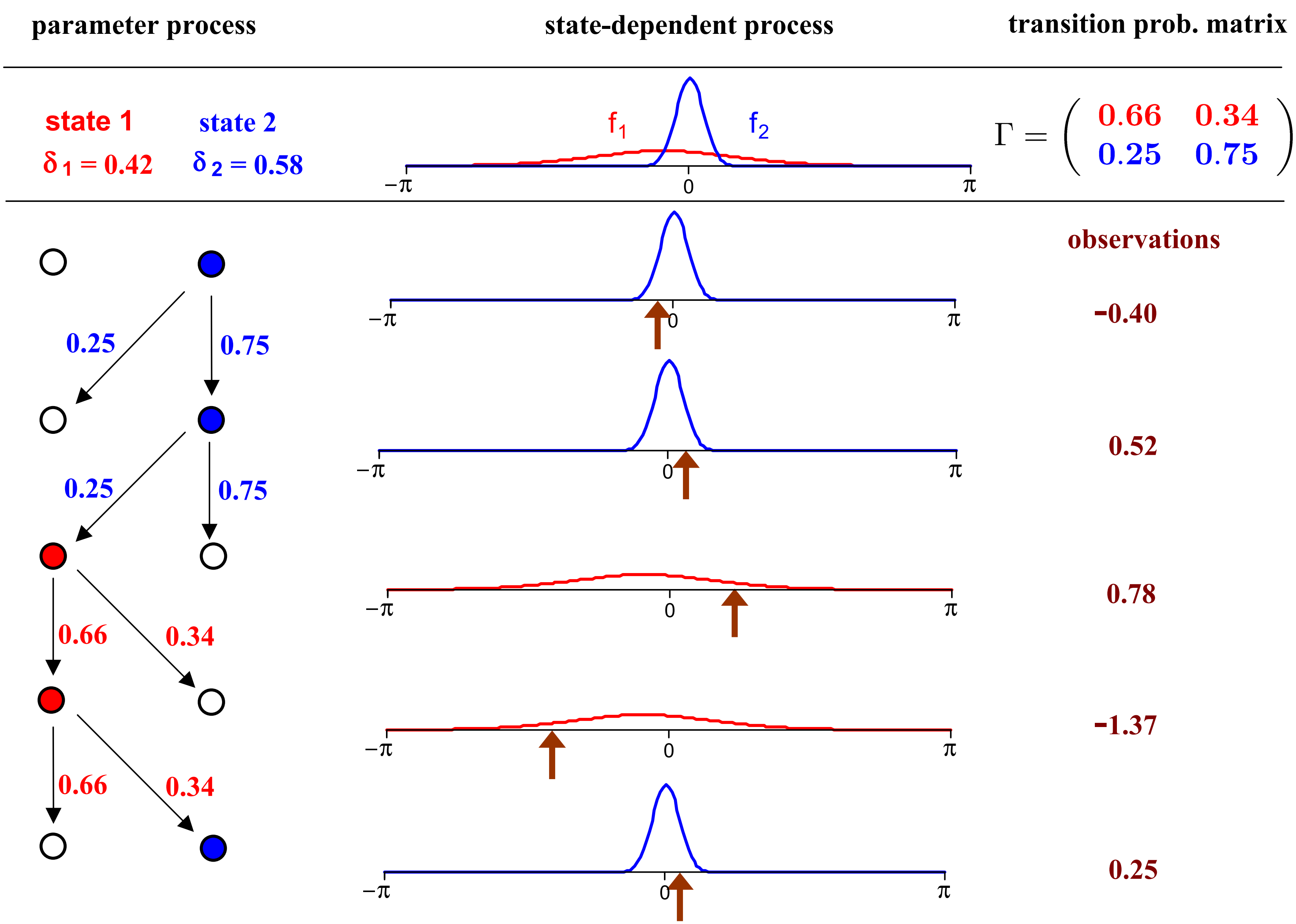
Definition of an HMM

$$\Pr(C_t \,|\, C^{(t-1)}) = \Pr(C_t \,|\, C_{t-1}) \quad \text{Markov property}$$

$$\Pr(A_t \,|\, A^{(t-1)}, C^{(t)}) = \Pr(A_t \,|\, C_t) \quad \text{Conditional independence}$$

all values up to time $t-1$.

Two-state von Mises-HMM



Two-state von Mises-HMM

hidden

observations
−0.40
0.52
0.78
−1.37
0.25

Parameters of a three-state stationary von Mises-HMM

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} \quad \& \quad \begin{matrix} \theta_1, \kappa_1 \\ \theta_2, \kappa_2 \\ \theta_3, \kappa_3 \end{matrix}$$

Stationary distribution:

$$\delta = \delta\Gamma$$

$$\text{m} - \text{state case}: \qquad m(m-1) \quad + \quad 2m \quad = m^2 + m$$

Properties: Convenient expressions for

- Marginal distributons ⇒ moments, likelihood
- Conditional distributons of the observations ⇒ residuals, forecasts
- Conditional distributons of the states ⇒ decoding, state prediction

The likelihood $\mathbf{L}_T = \underbrace{\delta\, \mathbf{B}_1\, \mathbf{B}_2 \cdots \mathbf{B}_t}_{\alpha_t} \underbrace{\mathbf{B}_{t+1} \cdots \mathbf{B}_T\, \mathbf{1}'}_{\beta_t'}$ with $\alpha_0 := \delta$, $\beta_T' := \mathbf{1}'$.

Forecast distribution

$$\Pr(A_{T+h} = a \mid A^{(T)} = a^{(T)}) = \frac{\alpha_T\, \Gamma^h\, \mathbf{P}(a)\, \mathbf{1}'}{\alpha_T \mathbf{1}'} = \sum_{i=1}^{m} \xi_i\, f_i(a)$$

Methods

- Parameter estimation — Baum-Welch or direct maximization,
- Standard errors, — e.g. parametric bootstrap,
- Model selection — classical and Bayesian,
- Model checking — quantile residuals,
- Global decoding — Viterbi algorithm.

Observations: $a_1 \quad a_2 \quad a_3 \quad \cdots \quad a_T$

Likelihood: $\delta \quad \Gamma\,\mathbf{P}(a_1) \quad \Gamma\,\mathbf{P}(a_2) \quad \Gamma\,\mathbf{P}(a_3) \quad \cdots \quad \Gamma\,\mathbf{P}(a_T) \quad \mathbf{1}'$

$$\Gamma\mathbf{P}(a) = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} \begin{pmatrix} f_1(a,\theta_1,\kappa_1) & 0 & 0 \\ 0 & f_2(a,\theta_2,\kappa_2) & 0 \\ 0 & 0 & f_3(a,\theta_3,\kappa_3) \end{pmatrix}$$

1. Baum-Welch Algorithm (EM)

Regard the hidden states, $C_1, C_2, ..., C_T$, as *missing observations.*

Apply the Expectation-Maximization (EM) algorithm.

2. Direct maximization of the likelihood function, $\mathbf{L}_T$

Use an algorithm to maximize the likelihood function directly.

Parameter constraints need to be respected.

Some issues — both cases

- Scaling is needed to avoid numerical underflow.
- These algorithms find a *local* maximum of the likelihood.

Baum-Welch Algorithm

⊕ Popular: Used more often than direct maximization.

⊕ Seems to be less sensitive to starting values[1].

⊕ Guaranteed increase in the likelihood at each iteration.

⊖ Needs a numerical "fix" to fit a stationary model. (δ is estimated separately.)

Direct maximization of the likelihood function

⊕ Faster convergence when approaching a maximum[1].

⊕ **Flexibility:** Easy to adapt for fitting new, or non-standard, models.

⊖ One has to take care of parameter constraints, e.g. via reparameterization.

[1] Berzel and Bulla (2006)

Conditional distributions of the unobserved states

$$\Pr(C_t = i \,|\, A^{(T)} = a^{(T)}) \times \mathbf{L}_T = \begin{cases} \alpha_T (\Gamma^{t-T})_{\bullet i} & t > T & \text{state prediction} \\ \alpha_T(i) & t = T & \text{filtering} \\ \alpha_t(i)\beta_t(i) & 1 \le t < T & \text{smoothing} \end{cases}$$

Notation: $B_{\bullet i}$ denotes the i*th* column of the matrix B.

Local decoding: the *a posteriori* most probable state at time t is

$$i_t^* \quad = \quad \underset{i \in \{1,\ldots,m\}}{\text{argmax}} \Pr(C_t = i \,|A^{(T)} = a^{(T)}), \quad t = 1, 2, \ldots, T$$

Global decoding: the *a posteriori* most probable *sequence of states*

$$(i_1^*, \ldots, i_T^*) \quad = \underset{i_1,\ldots,i_T \in \{1,2,\ldots,m\}}{\text{argmax}} \Pr(C_1 = i_1, \ldots, C_T = i_T \,|\, A^{(T)} = a^{(T)})$$

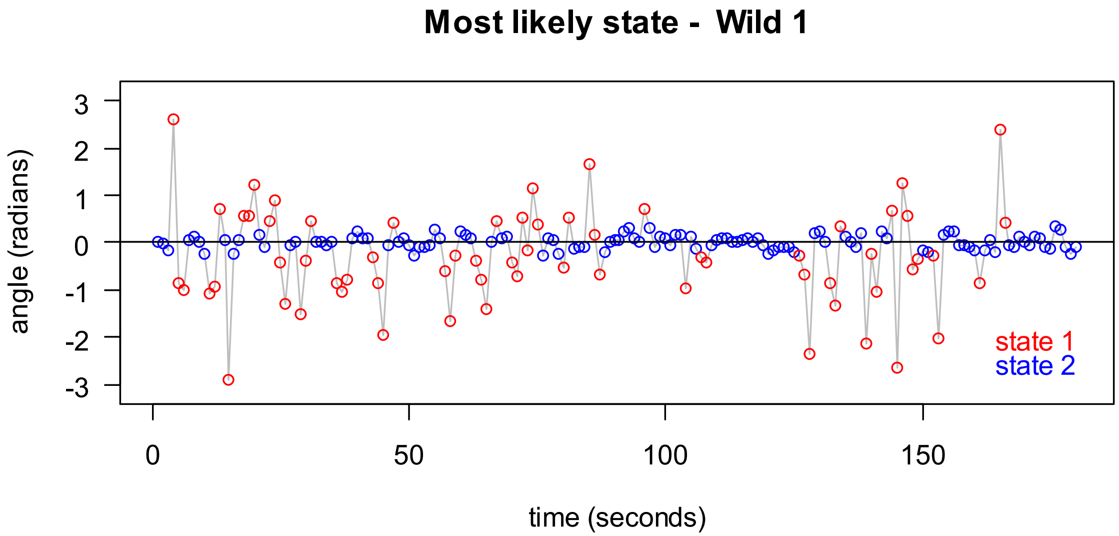
Computed using the Viterbi algorithm (dynamic programming)

Parameter estimates and local decoding

Estimates for wild 1

$$\hat{\Gamma} = \begin{pmatrix} 0.66 & 0.34 \\ 0.25 & 0.75 \end{pmatrix}$$

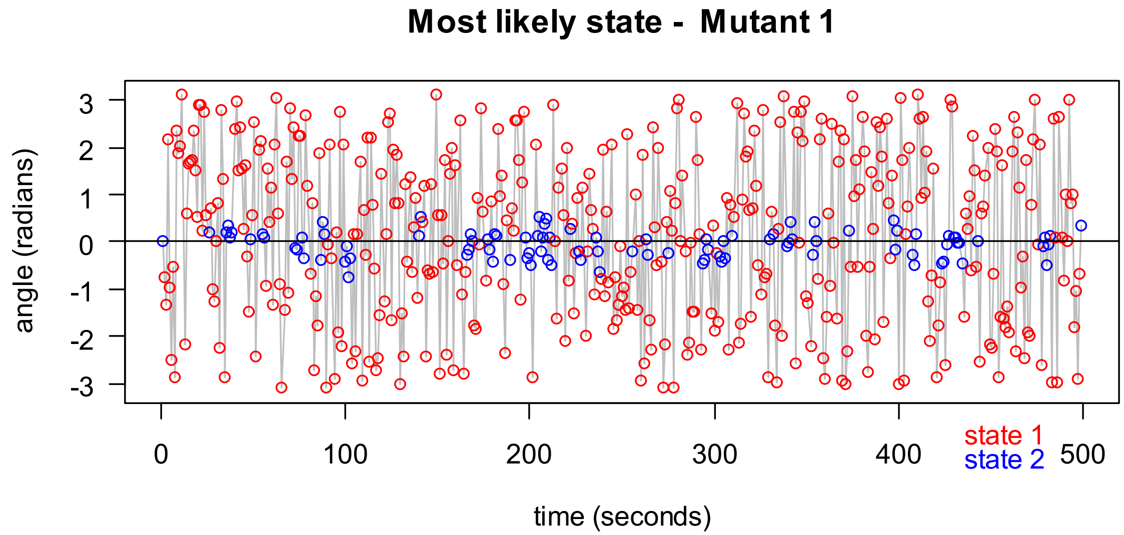
$\hat{\delta}$	$\hat{\theta}$	$\hat{\kappa}$
0.42	−0.28	1.65
0.58	0.02	41.96



Estimates for mutant 1

$$\hat{\Gamma} = \begin{pmatrix} 0.86 & 0.14 \\ 0.54 & 0.46 \end{pmatrix}$$

$\hat{\delta}$	$\hat{\theta}$	$\hat{\kappa}$
0.79	1.91	0.12
0.21	−0.03	7.68



Defining appropriate residuals for hidden Markov models

- The conditional distribution of A_t given $a_1 \ldots, a_{t-1}, a_{t+1}, \ldots, a_T$ changes for every t.

- So does that of A_t given $a_1, \ldots, a_{t-1}$.

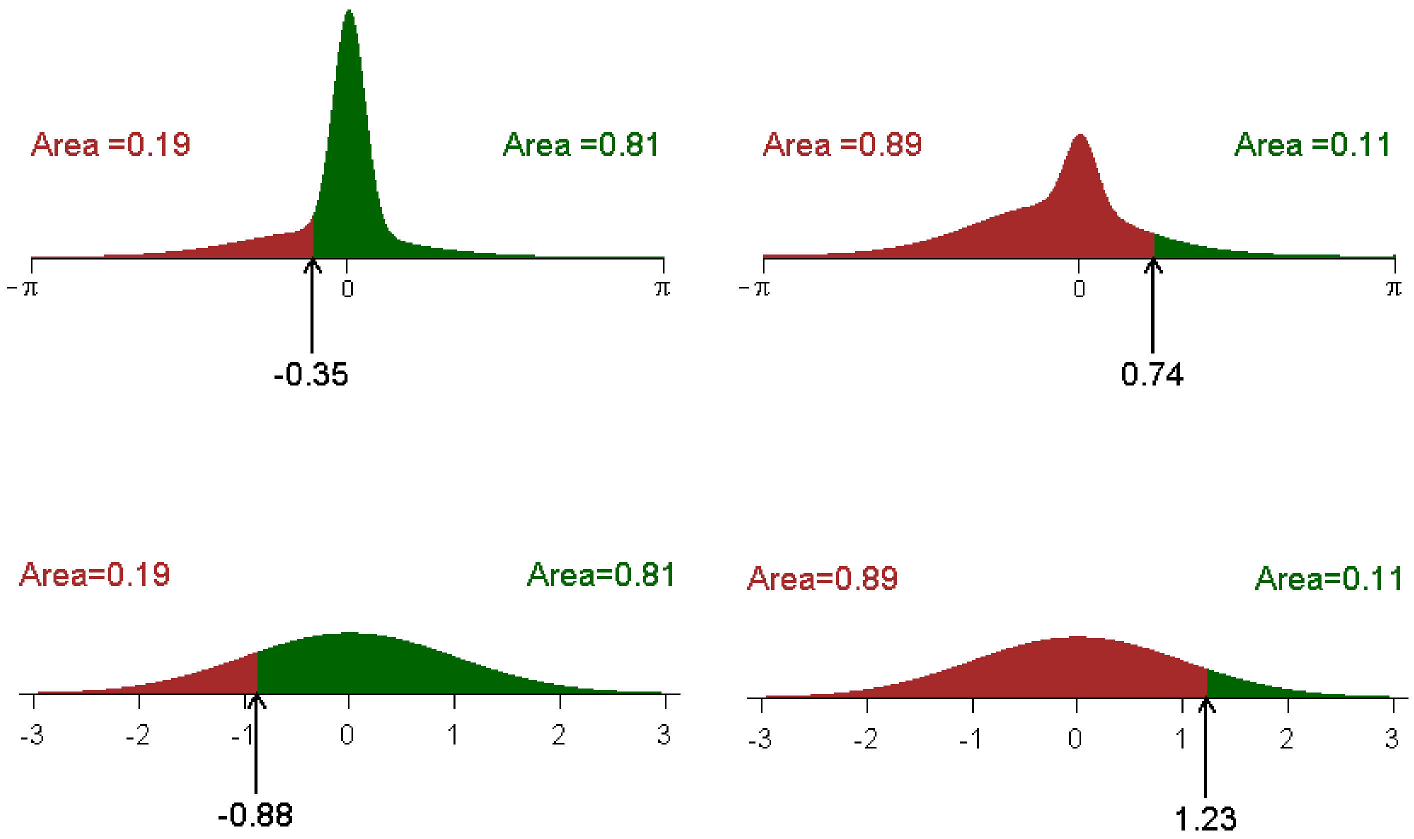
- Ordinary residuals: $e_t = a_t -$ *"conditional expectation"* each have quite different distributions.

- We need other quantities to construct residual plots, qqplots, etc., for example quantile residuals.

- Forecast normal quantile residual: $r_t = \Phi^{-1}(\tilde{F}(a_t))$,

  where $\tilde{F}_t(a) = \int_{-\pi}^{a} \tilde{f}_t(x)\,dx \quad t = 1, 2, \ldots, T.$

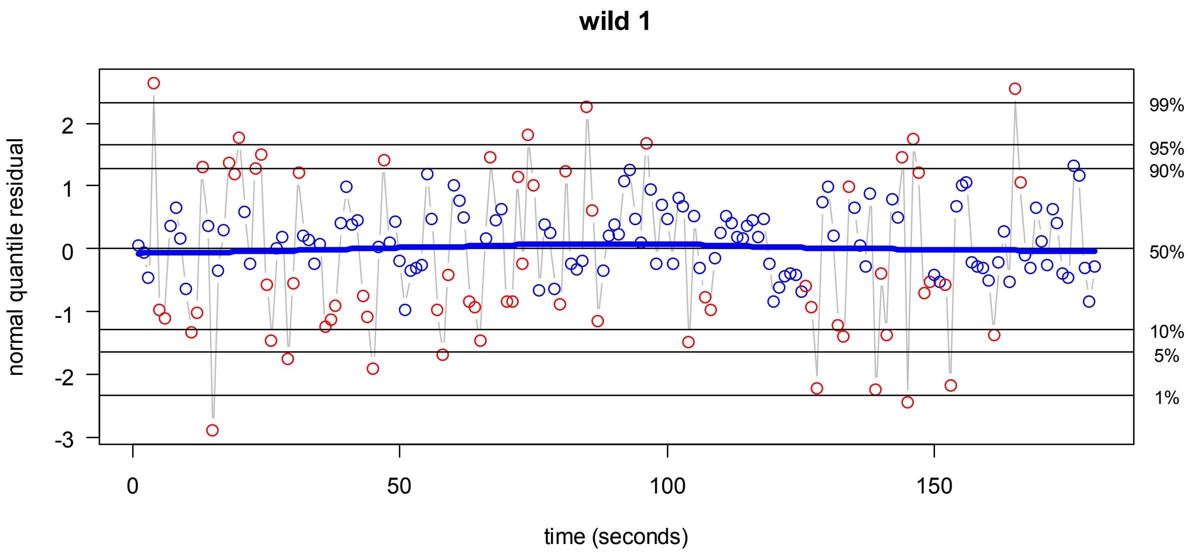
Two observations and their forecast distributions

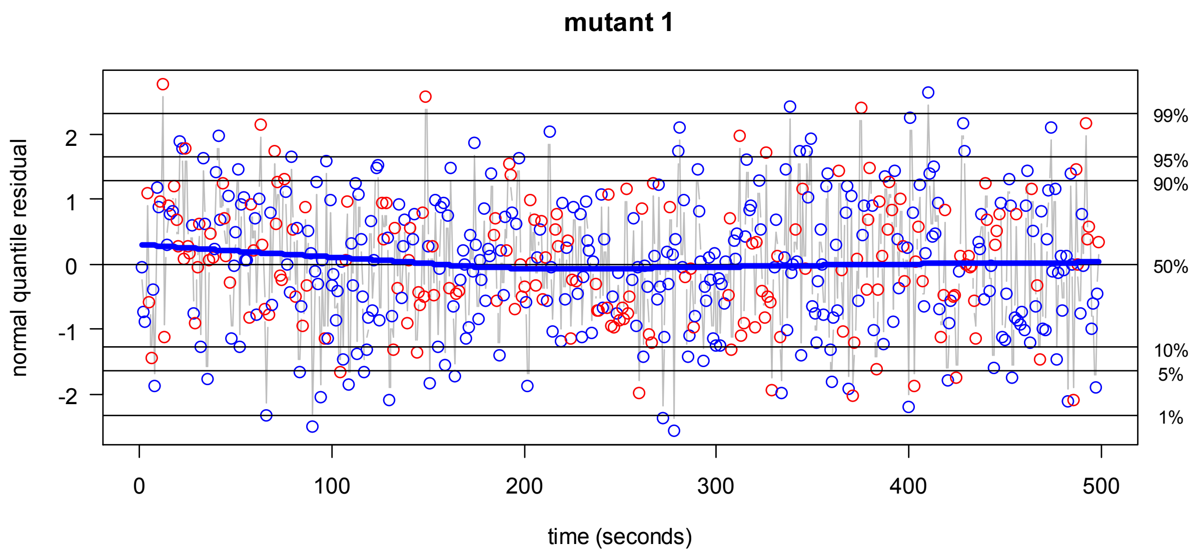


Standard normal

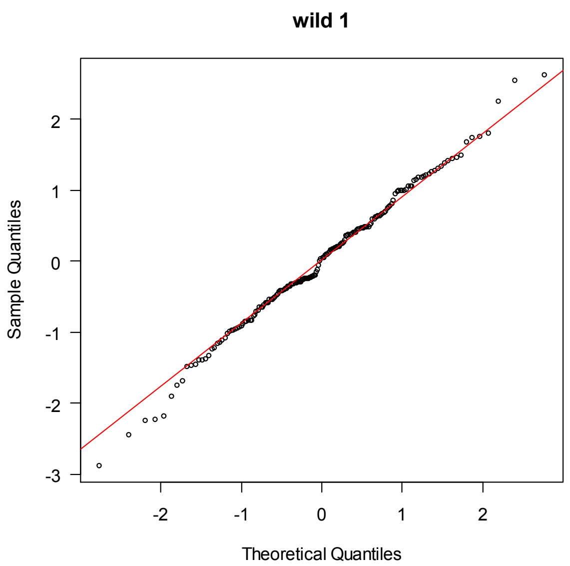
Residuals:

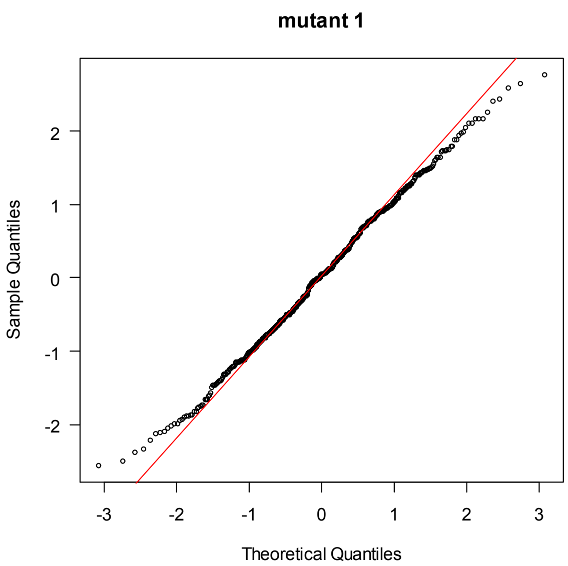
Forecast quantile residuals and a non-parametric smooth

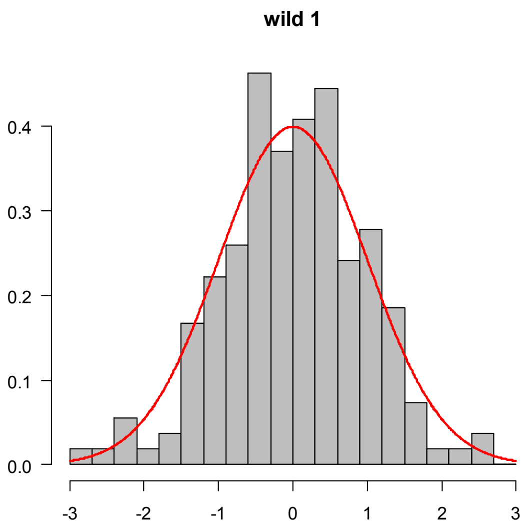


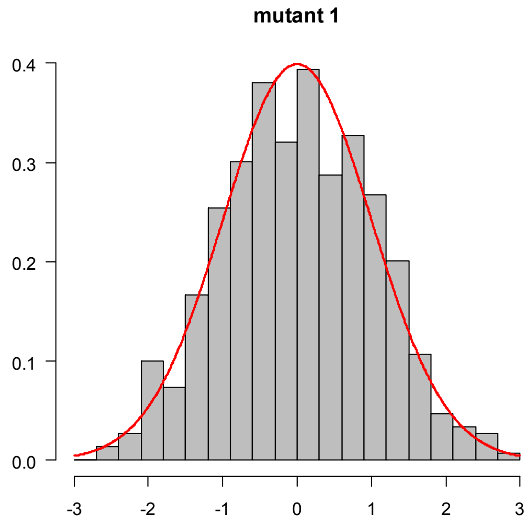


Quantile residuals







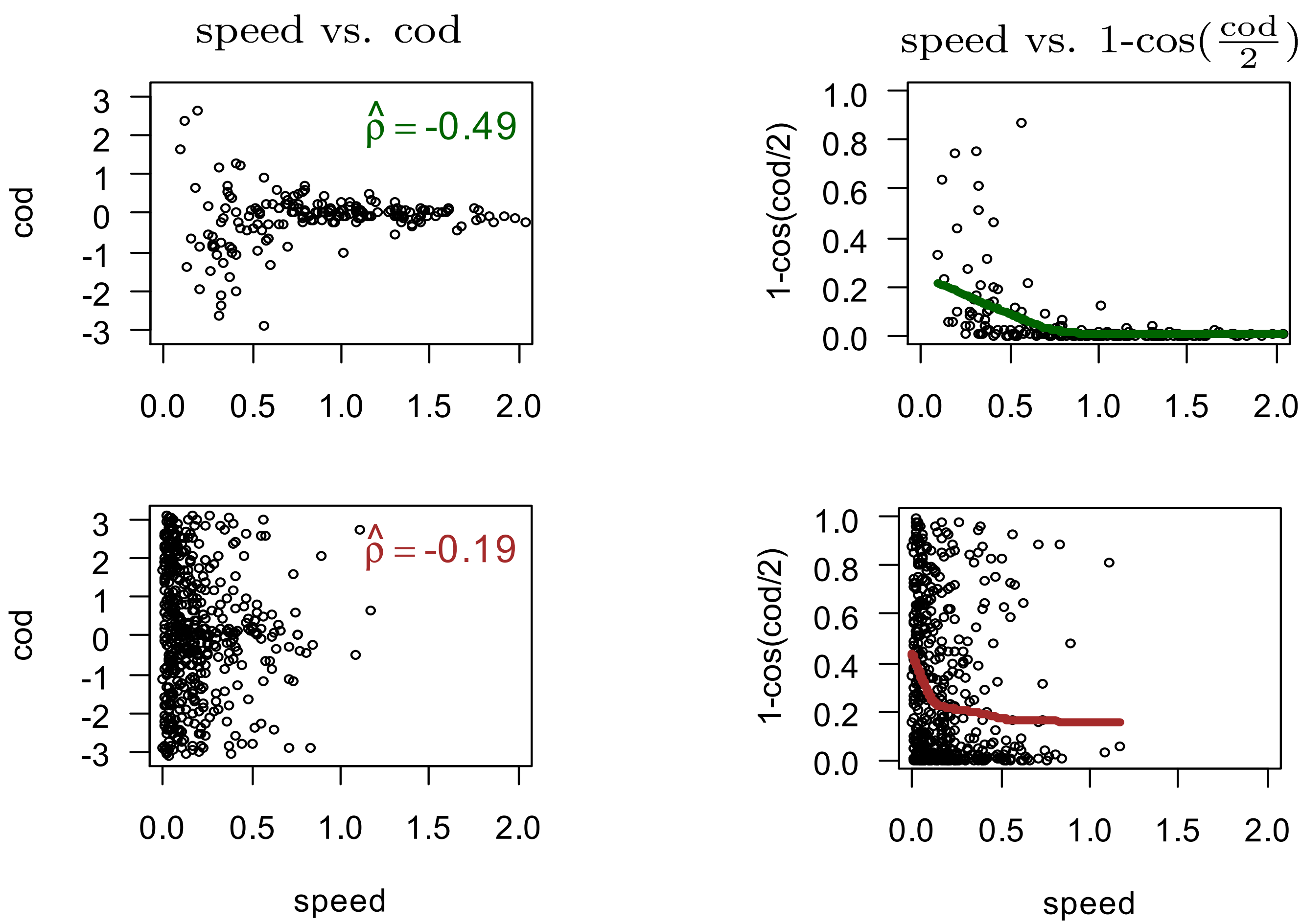


Modelling speed and change of direction.

Speed and change of direction are negatively correlated[1].

Wild 1

Mutant 1



[1]For correlation between linear and circular variables see Mardia (1976).

Model

Bivariate HMM assuming contemporaneous conditional independence:

$$f(a_t, s_t \,|\, C_t = i) = f(a_t \,|\, C_t = i)\, f(s_t \,|\, C_t = i)$$

The state-dependent distributions

von Mises: $f(a) = \frac{1}{2\pi I_0(\kappa)}\, e^{\kappa \cos(a-\theta)}$ $\quad -\pi \leq a < \pi$

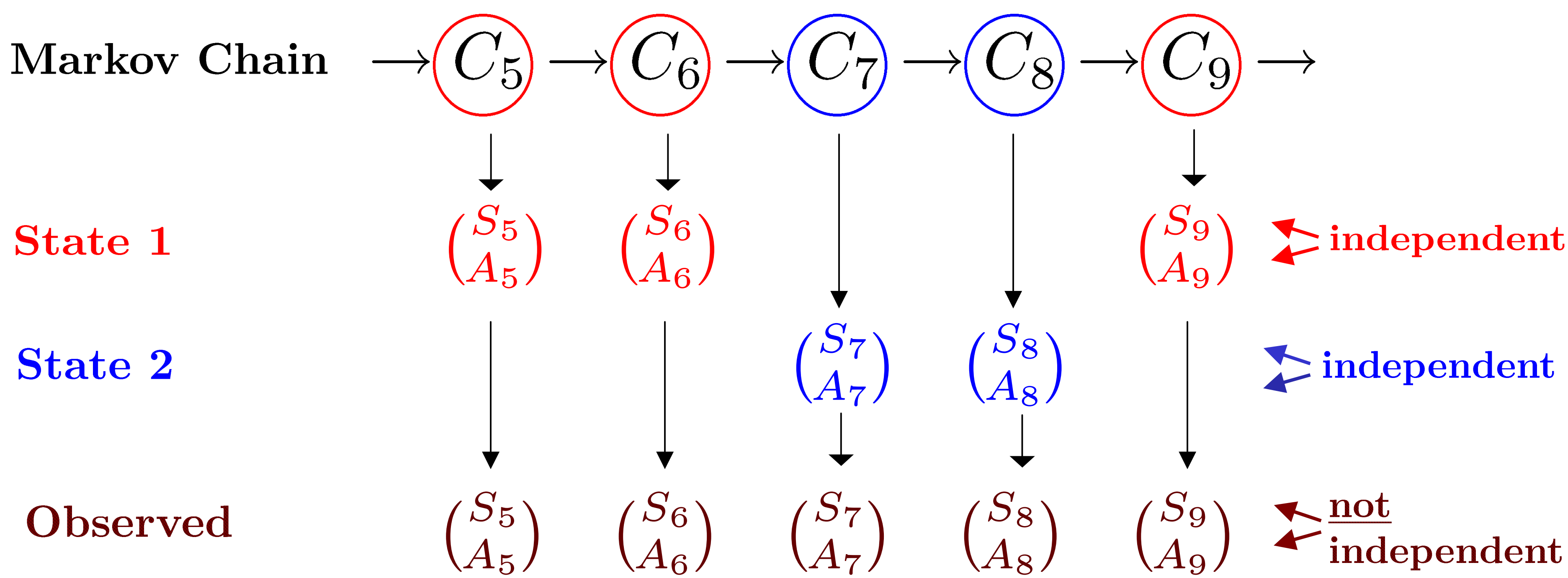
Gamma: $f(s) = \frac{\gamma^\nu}{\Gamma(\nu)}\, s^{\nu-1} e^{-s\lambda}$ $\quad s \geq 0$

fit the series for most, but not all, individuals quite well.

Note: Contemporaneous conditional independence $\not\Rightarrow$ independence.

Contemporaneous conditional independence

Contemporaneous conditional independence $\not\Rightarrow$ independence.



- Markov chain ⇒ **serial dependence**,
- Unequal state–dependent distributions ⇒ **contemporaneous dependence**.

Selected estimates for all larvae:

State 1

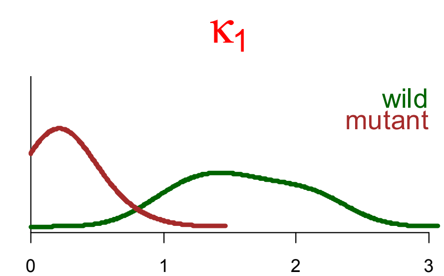
turning/head–swinging
large cods, low speed

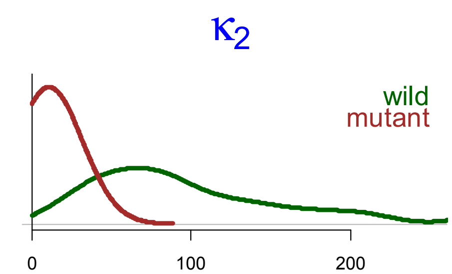
State 2

speedy linear locomotion
Small cods, high Speed

Von Mises parameter κ

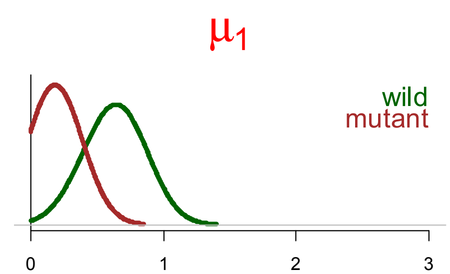
wild *less* dispersed in both states

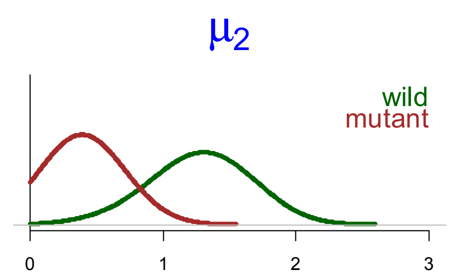




Gamma distribution mean

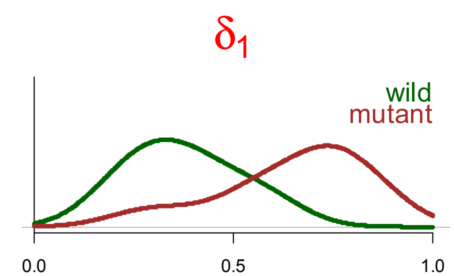
wild *faster* in both states

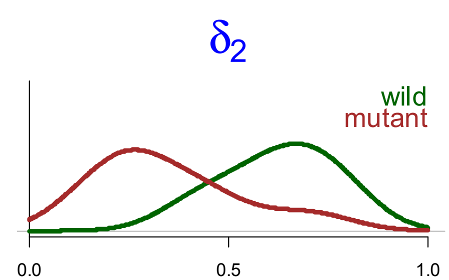




Stationary dist. of Markov Chain

wild spend *less* time in state 1





Wind direction at Koeberg

Data:	Average hourly wind direction
Period:	01.05.1985 – 30.04.1989
Length:	35 064 observations
Observations:	One of 16 compass directions
Code:	N=1, NNE=2, ..., NNW=16
Objective:	One-hour-ahead forecast of direction

Models for the hourly series

0. First-order Markov chain — baseline model
1. Categorical–HMM
2. *Seasonal* categorical–HMM

Models for the daily series

0. First-order Markov chain — baseline model
1. Categorical–HMM
2. Circular–valued HMM

Models for change in direction

1. Von-Mises–HMM
2. Von-Mises–HMM with wind speed as covariate - version 1
3. Von-Mises–HMM with wind speed as covariate - version 2

Categorical–HMM and its likelihood function

Observations: $a_t = (a_{t1}, a_{t2}, \ldots, a_{t16})$

$$\text{where } a_{tj} = \begin{cases} 1 & \text{if the wind direction is } j \text{ at time } t, \\ 0 & \text{if it isn't.} \end{cases}$$

Example $a_t = (1, 0, 0, \ldots, 0)$ indicates $j = 1$ (North) at time t.

State–dependent distribution: $\Pr(\text{direction } j \mid \text{state } i) = \pi_{ji}$

Likelihood: $\delta\Gamma\, \mathrm{P}(a_1)\Gamma\, \mathrm{P}(a_2)\Gamma\, \mathrm{P}(a_3) \cdots \Gamma\, \mathrm{P}(a_T)\mathbf{1}'$

where $\mathrm{P}(a_t)$ is a diagonal matrix with i–th entry

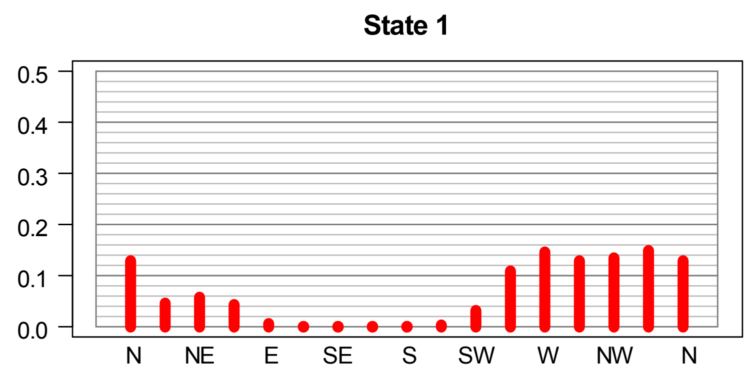
$$\Pr(A_t = a_t | \, C_t = i) = \pi_{1\,i}^{a_{t1}} \; \pi_{2\,i}^{a_{t2}} \; \cdots \; \pi_{16\,i}^{a_{t16}}$$

Estimates for two-state categorical–HMM

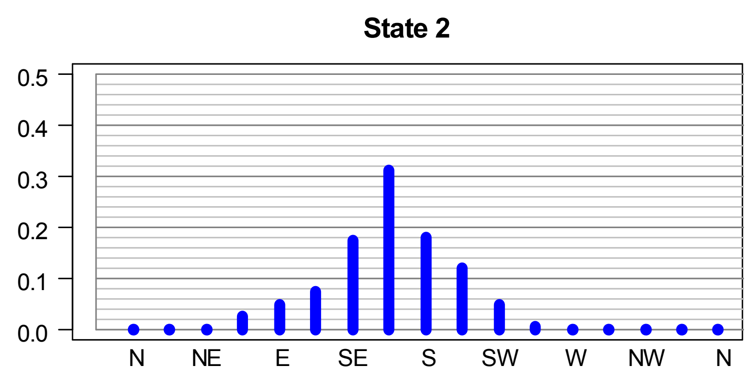
State–dependent model: $\Pr(A_t = j \mid C_t = i) = \begin{cases} \pi_{j1}, & \text{for } i = 1 \\ \pi_{j2}, & \text{for } i = 2 \end{cases}$

$$\hat{\Gamma} = \begin{pmatrix} 0.964 & 0.036 \\ 0.031 & 0.969 \end{pmatrix}$$

$$\hat{\delta}' = \begin{pmatrix} 0.462 \\ 0.538 \end{pmatrix}$$

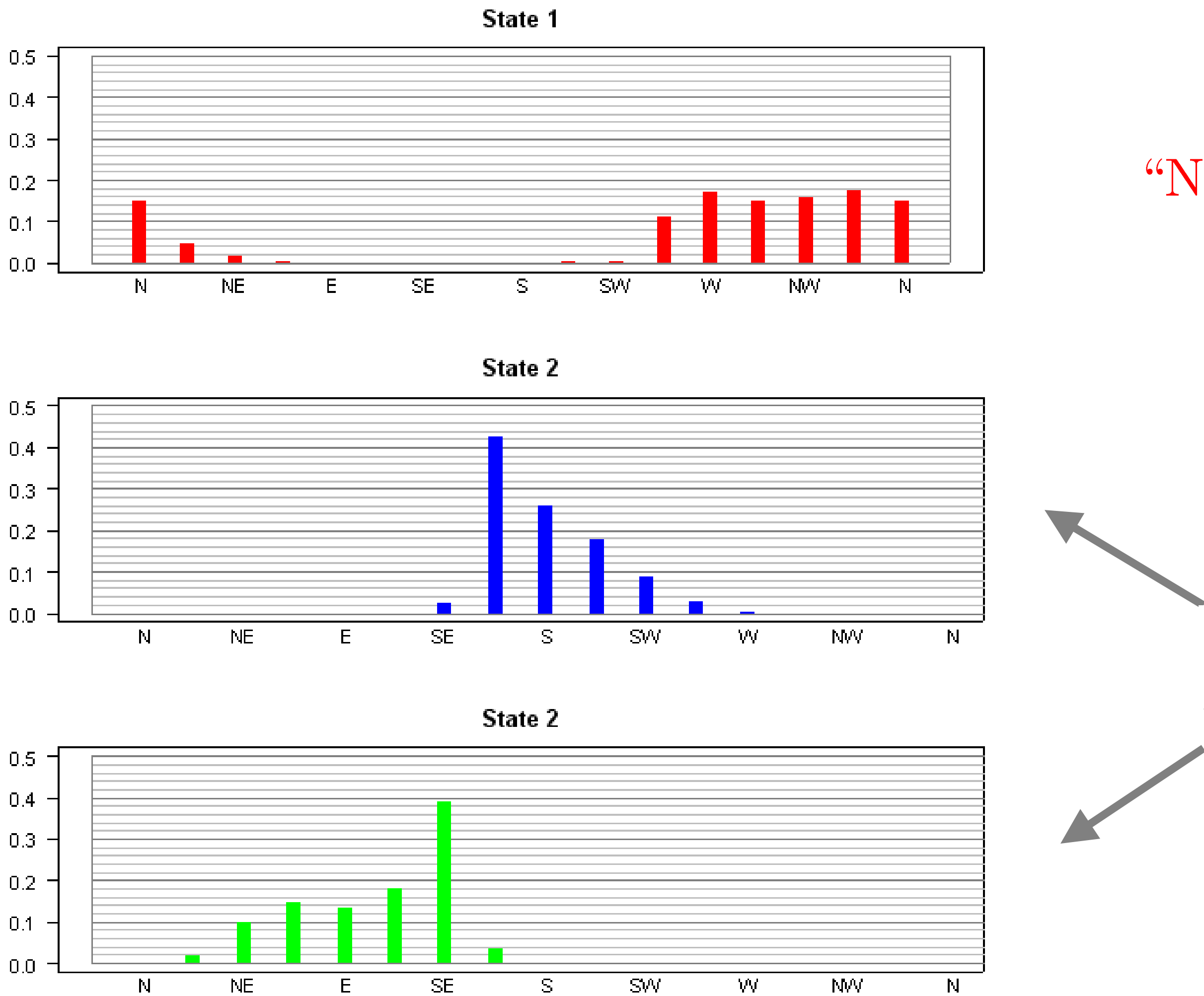


"North–westerly"



"South–easterly"

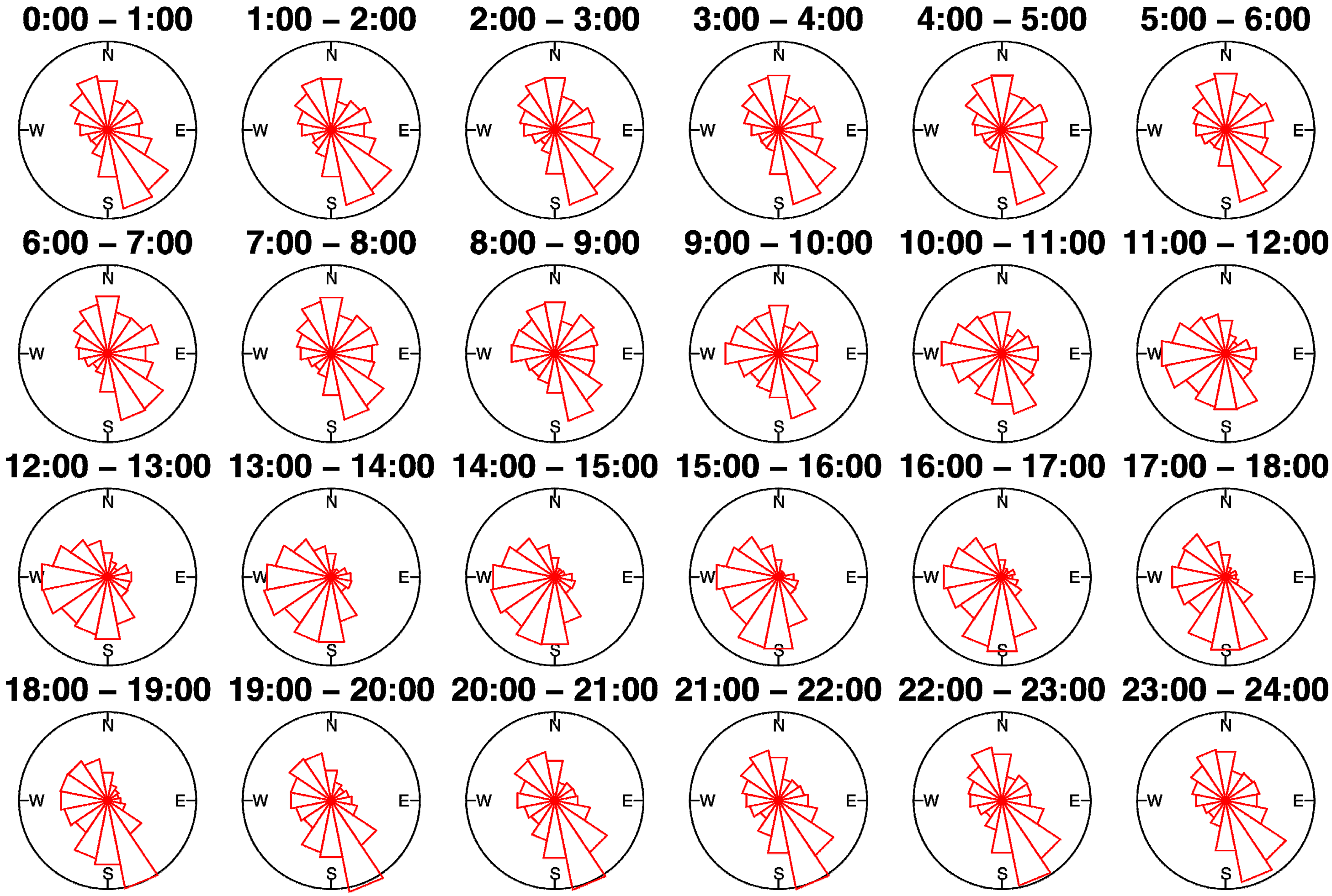
Three-state model — state-dependent distributions



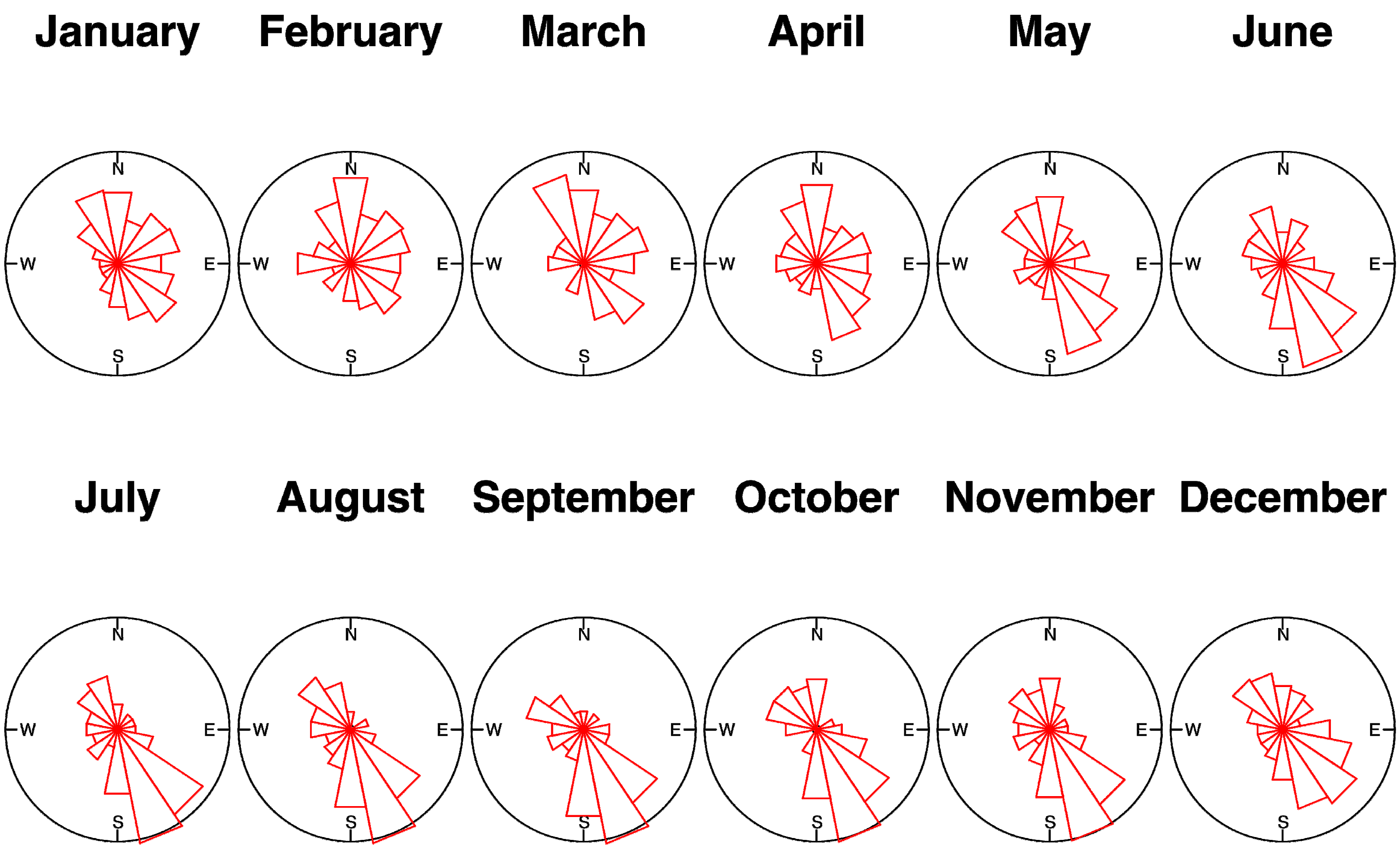
"North–westerly"

"South–easterly" is split in two.

Wind direction by time of day



Wind direction by month (23:00–24:00)



Seasonal **categorical–HMM**

State–dependent model:

$$\Pr(A_t = j \mid C_t = i) = \begin{cases} \pi_{j1}, & \text{for } i = 1 \\ \pi_{j2}, & \text{for } i = 2 \end{cases}$$

Transition probabilities are functions of **time**.

$$\Gamma = \begin{pmatrix} \gamma_{11}(t) & \gamma_{12}(t) \\ \gamma_{21}(t) & \gamma_{22}(t) \end{pmatrix}$$

$$\text{logit}(\gamma_{12}(t)) = a_1 + b_1 \cos\left(\frac{2\pi t}{24}\right) + c_1 \sin\left(\frac{2\pi t}{24}\right) + d_1 \cos\left(\frac{2\pi t}{8766}\right) + e_1 \sin\left(\frac{2\pi t}{8766}\right)$$

$$\text{logit}(\gamma_{21}(t)) = a_2 + \underbrace{b_2 \cos\left(\frac{2\pi t}{24}\right) + c_2 \sin\left(\frac{2\pi t}{24}\right)}_{\text{daily cycle}} + \underbrace{d_2 \cos\left(\frac{2\pi t}{8766}\right) + e_2 \sin\left(\frac{2\pi t}{8766}\right)}_{\text{annual cycle}}$$

Seasonal categorical-HMM — estimates

State–dependent model: $\Pr(X_t = j \mid C_t = i) = \begin{cases} \pi_{j1}, & \text{for } i = 1 \\ \pi_{j2}, & \text{for } i = 2 \end{cases}$

Parameters of $\Gamma(t)$

	$i = 1$	$i = 2$
$\hat{a}_i$	-3.349	-3.523
$\hat{b}_i$	0.197	-0.272
$\hat{c}_i$	-0.695	0.801
$\hat{d}_i$	-0.208	0.082
$\hat{e}_i$	-0.401	-0.089

General pattern is very similar to that of the simple two–state HMM.

j	Direction	π_{j1}	π_{j2}
1	N	0.127	0.000
2	NNE	0.047	0.000
3	NE	0.057	0.002
4	ENE	0.027	0.040
5	E	0.004	0.052
6	ESE	0.001	0.076
7	SE	0.001	0.179
8	SSE	0.000	0.317
9	S	0.001	0.183
10	SSW	0.007	0.121
11	SW	0.059	0.026
12	WSW	0.114	0.003
13	W	0.145	0.000
14	WNW	0.128	0.000
15	NW	0.135	0.000
16	NNW	0.147	0.000

Model selection criteria

model	#(pars)	-log(lk)/1000	AIC/1000	BIC/1000
Markov chain	240	48	97	99
2-state HMM	32	76	152	152
3-state HMM	51	70	139	140
2-state seasonal HMM	40	76	151	152

- The HMM models don't even come close to beating the first–order Markov chain. (Sad but true.)
- Reason: The HMMs don't take previous direction into account.

Circular-valued HMM

Regard the observations as *interval-censored* von Mises random variables.

$$\Pr(\text{direction} = j) = \pi_j = \int_{\frac{2\pi(j+0.5)}{16}}^{\frac{2\pi(j-0.5)}{16}} f_{\text{vM}}(a)\, da\,, \quad j = 1, \ldots 16.$$

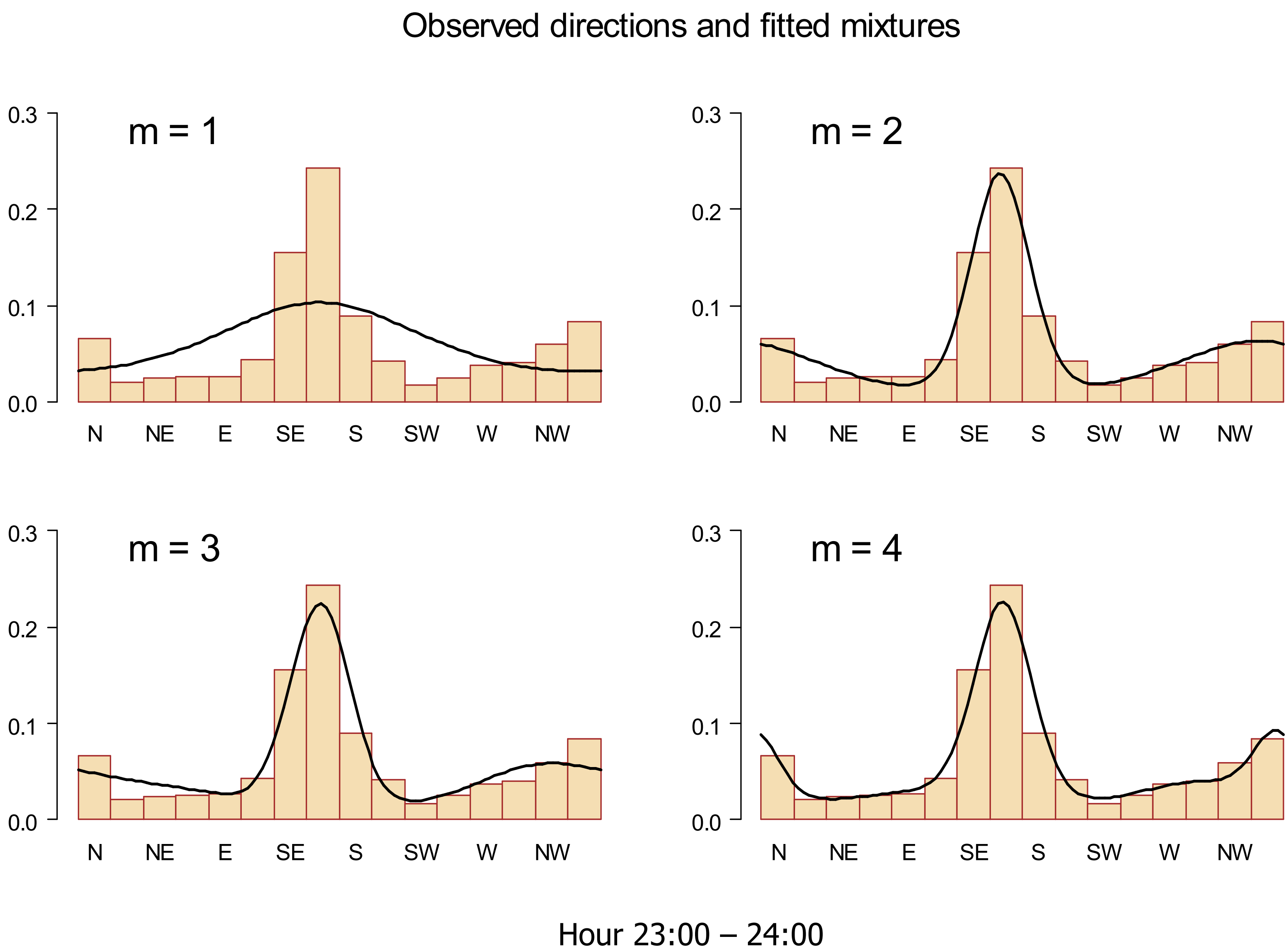
Observed values $\sim$ Multinomial$(1, \pi_1, \pi_2, \ldots, \pi_{16})$

The 16 values π_j are determined by the 2 parameters θ and κ.

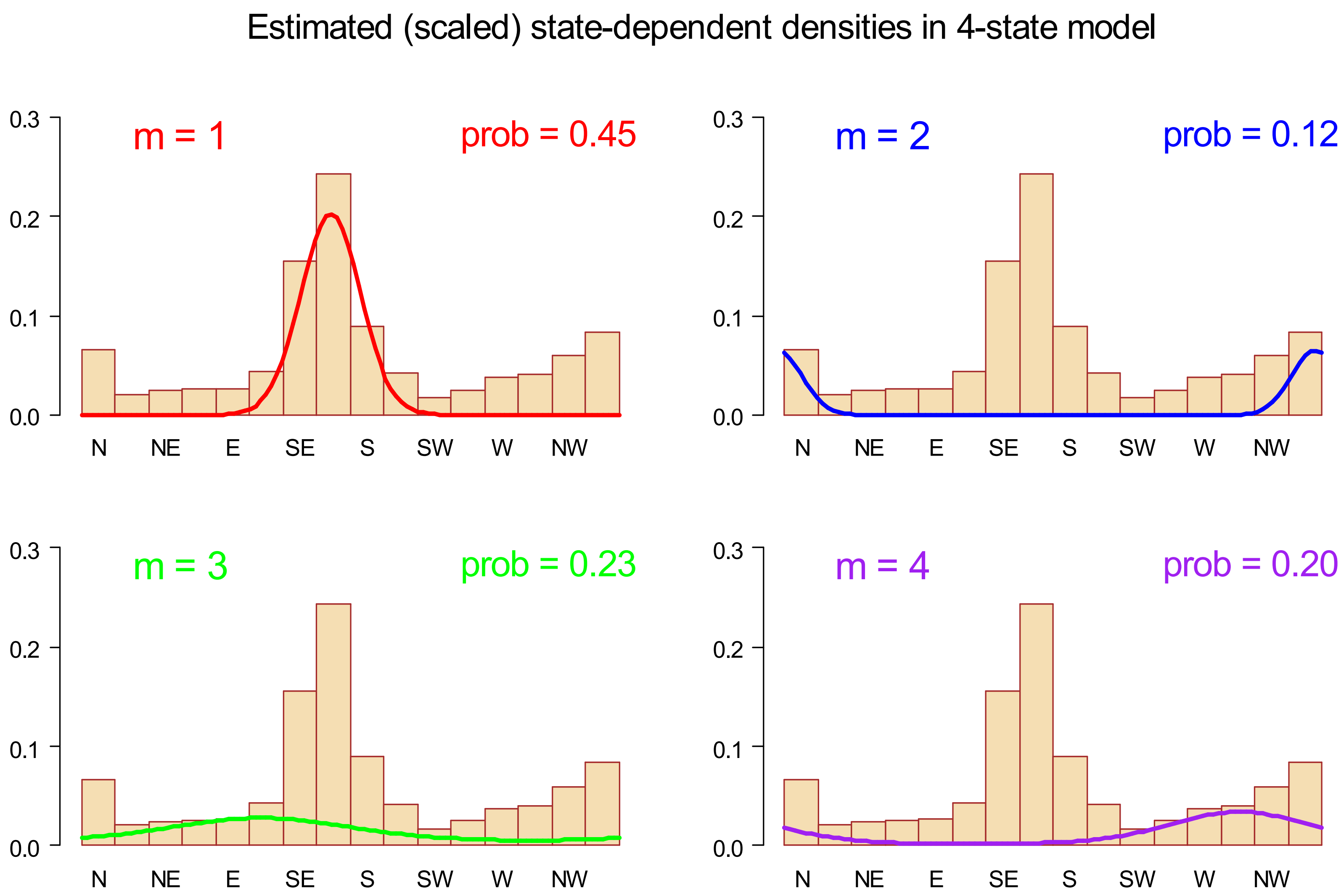
Two-state von Mises(θ_i, κ_i)-HMM

State 1: $\pi_{1\,1}, \pi_{2\,1}, \ldots, \pi_{16\,1}$ are determined by θ_1, κ_1
State 2: $\pi_{1\,2}, \pi_{2,2}, \ldots, \pi_{16\,2}$ are determined by θ_2, κ_2

Wind direction at Koeberg - daily series



Wind direction at Koeberg - daily series



Hour 23:00 – 24:00

Daily series

Average direction over the hour 23:00 – 24:00 (1461 observations).

model	#(pars)	-log(lk)/10	AIC/10	BIC/10
1-state von Mises-HMM	2	393	787	788
2-state von Mises-HMM	6	361	723	726
3-state von Mises-HMM	12	354	710	716
4-state von Mises-HMM	20	349	701	712
2-state multinomial-HMM	32	346	699	716
Saturated Markov chain	240	329	707	833

The von Misess-HMM is not much better here. (Nice try, but no cigar.)

General point:

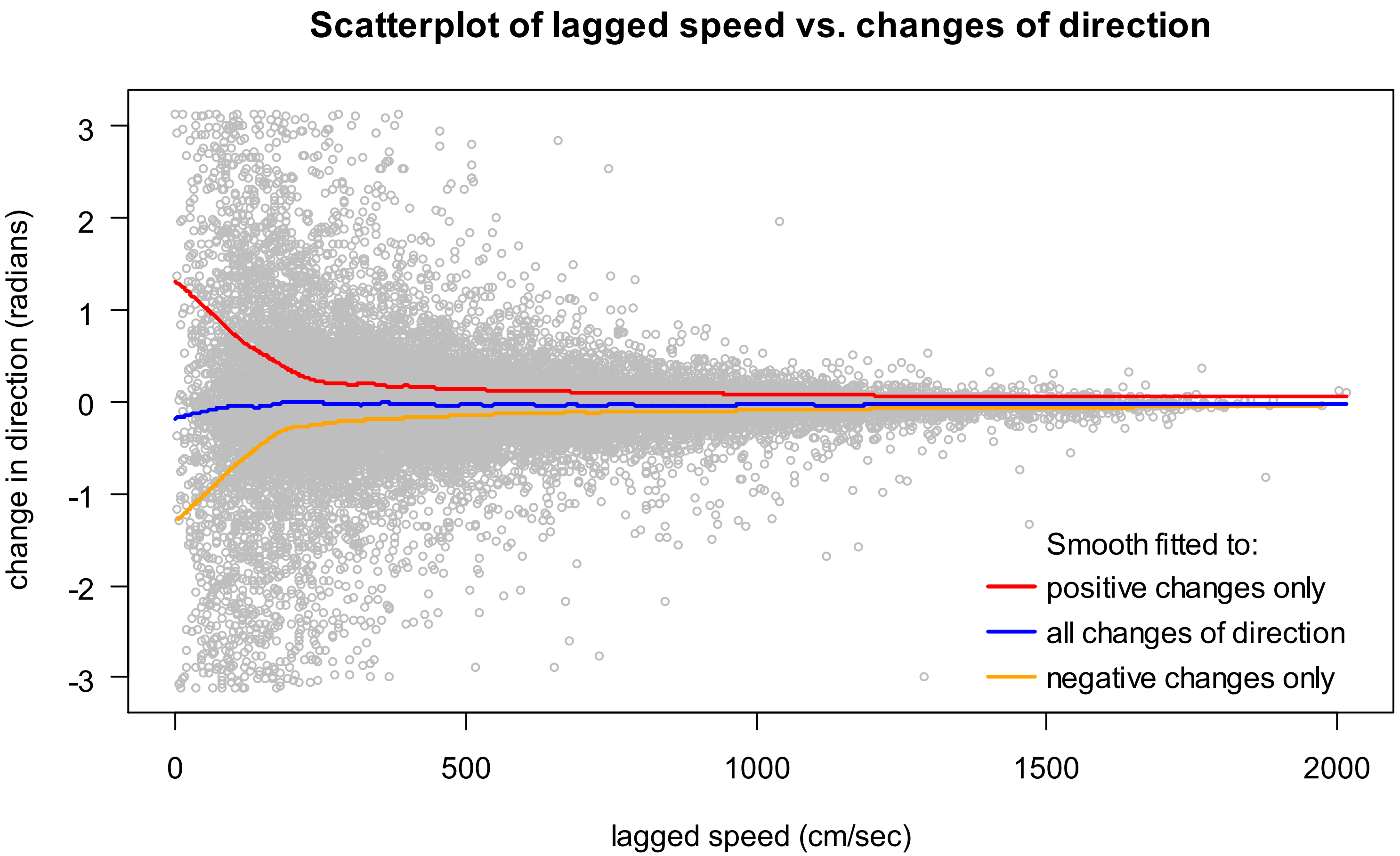
This example illustrates that one can fit HMMs when

- the observations are interval-censored,
- some observations are interval-censored, some are not,
- some observations are missing (at random) — extreme censoring!

HMM with speed as covariate

Observations: hourly speed (cm/sec) and direction (degrees).

New objective: model the change of direction.



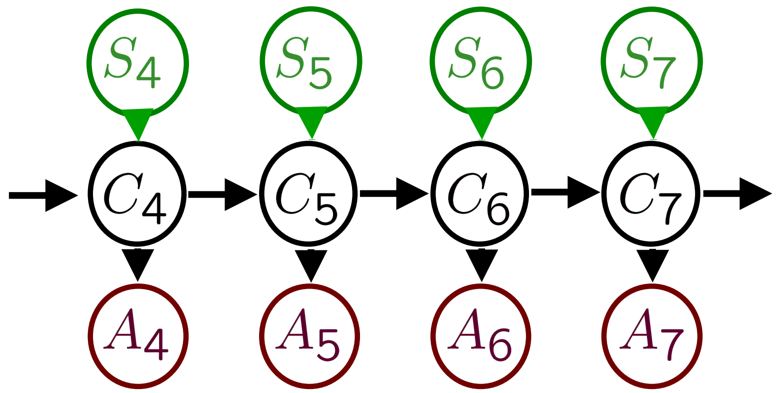
Version 1: Speed affects the Markov Chain

High speed makes the transitions between states less likely.

Covariate $\left(\sqrt{\text{speed}}\right)$

State process

State-dependent process (change of direction)



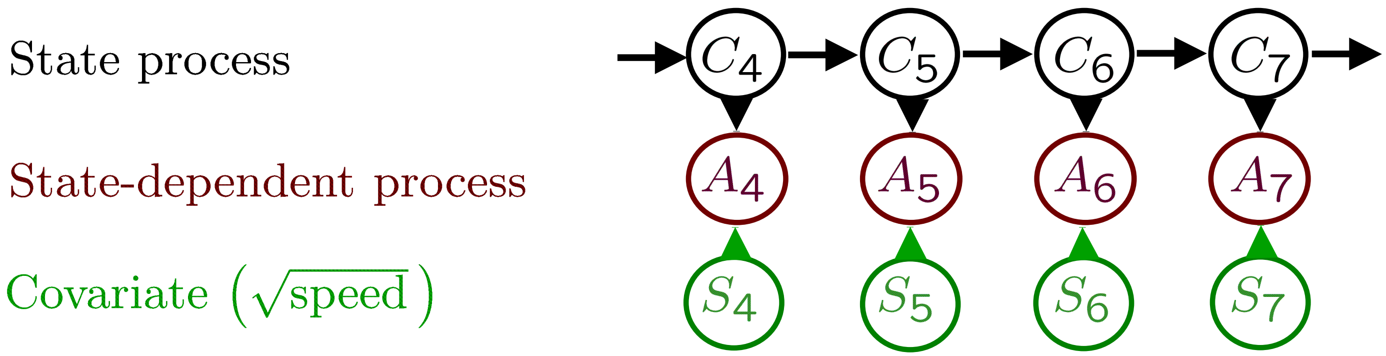
Model: von Mises-HMM with transition probability matrix:

$$\Gamma(s_{t-1}) = \begin{pmatrix} \gamma_{11}(s_{t-1}) & \gamma_{12}(s_{t-1}) \\ \gamma_{21}(s_{t-1}) & \gamma_{22}(s_{t-1}) \end{pmatrix}$$

$$\text{with } \gamma_{ij}(s) = \frac{e^{\tau_{ij}}}{\sum_{k=1}^{m} e^{\tau_{ik}}} \quad \text{and} \quad \log \tau_{ii} = \eta_i \sqrt{s}\,, \qquad i,j = 1,2,\ldots,m.$$

Version 2: Speed affects the von-Mises dispersion parameter

High speed reduces the dispersion.



Model: von Mises-HMM with speed-dependent dispersion paremeters

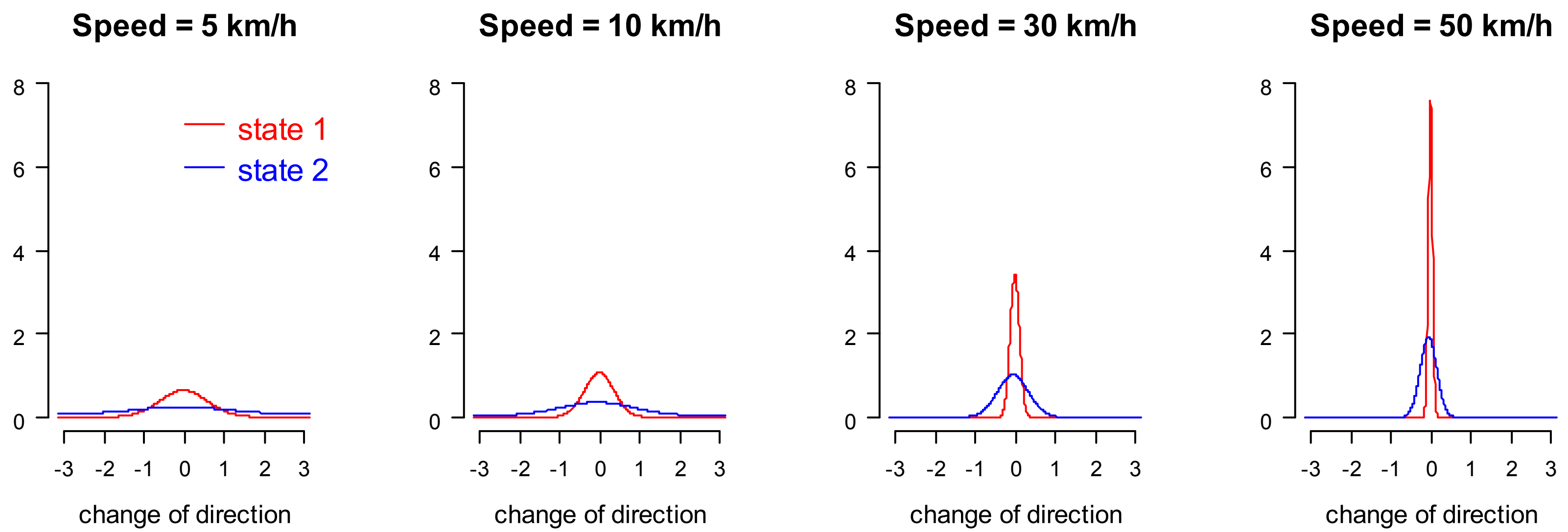
State 1: $A_t \sim \text{vM}(\theta_1, \kappa_1), \quad \kappa_1 = e^{\alpha_{01} + \alpha_{11}\sqrt{s_{t-1}}}$
State 2: $A_t \sim \text{vM}(\theta_2, \kappa_2), \quad \kappa_2 = e^{\alpha_{02} + \alpha_{12}\sqrt{s_{t-1}}}$

Model comparison

model	covariate	m	#(pars)	-log(lk)/100	AIC/100	BIC/100
1	none	1	2	218	436	437
		2	6	86	172	173
		3	12	70	140	141
		4	20	68	136	137
2	Speed affects the	2	8	68	136	136
	Markov chain	3	15	56	112	113
		4	25	54	108	110
3	Speed affects the	1	3	104	209	209
	dispersion parameter	2	5	52	104	104
		3	7	50	101	101

- Models for change of direction lead to *much more accurate* one-hour-ahead forecasts than models for direction.
- Forecasts improve if one uses speed as a covariate.
- Model 3 has a nasty likelihood surface — estimation is tricky!

Model 3: State-dependent von Mises densities for four (lagged) wind speeds



Using speed as a covariate improves the forecasts substantially.

Circular-valued HMMs

- The circular nature of the data presents no problems.
- Covariates can be included in different ways.
- Censored and missing observations can be dealt with precisely.
- They can model – multivariate series,
  – bivariate linear-valued and circular-valued series,
  – series with multimodal marginal distributions.
- They are satisfyingly flexible.

Of course, like any other models, they don't fit everything!

Iloilo aligatoh gozaimaschta!