Biometrical Journal **43** (2001) 7, 863–879

Finite Sample Properties of the Maximum Likelihood Estimator and of Likelihood Ratio Tests in Hidden Markov Models

S. MICHALEK, M. WAGNER, and J. TIMMER[1]

FDM-Center for Data Analysis, University of Freiburg, Germany

W. VACH

Department of Statistics and Demography, Odense University, Denmark

Summary

Hidden Markov models were successfully applied in various fields of time series analysis, especially for analyzing ion channel recordings. The maximum likelihood estimator (MLE) has recently been proven to be asymptotically normally distributed. Here, we investigate finite sample properties of the MLE and of different types of likelihood ratio tests (LRTs) by means of simulation studies. The MLE is shown to reach the asymptotic behavior within sample sizes that are common for various applications. Thus, reliable estimates and confidence intervals can be obtained. We give an approximative scaling function for the estimation error for finite samples, and investigate the power of different LRTs suitable for applications to ion channels, including tests for superimposed hidden Markov processes. Our results are applied to physiological sodium channel data.

Key words: Hidden Markov model; Ion channel analysis; Parameter estimation; Maximum likelihood estimator; Likelihood ratio test; Finite sample properties; Asymptotic normality; Superposition of Markov processes; Power of test; Sodium channel.

1. Introduction

Hidden Markov models (HMM) were successfully applied in various fields of statistical time series analysis, e.g. in ion channel analysis (CHUNG et al., 1990; FREDKIN and RICE, 1992; BECKER et al., 1994; MILBURN et al. 1995; CHUNG and KENNEDY, 1996; MICHALEK et al. 1999), protein and nucleic acid sequence analysis (KROGH et al., 1994; DURBIN et al., 1998), communication technology (ELLIOTT et al., 1995), econometrics (HAMILTON, 1990) and speech recognition (RABINER, 1989).

In biological applications, it is often difficult to keep the experimental conditions fixed. Therefore, there is often a limited amount of data available from these experiments. Although maximum likelihood parameter estimation is commonly

[1] Phone +49 761 203–5829, Fax –7700, E-mail: Jens.Timmer@fdm.uni-freiburg.de

 0323-3847/01/0711-0863 $ 17.50+.50/0

used, the applicability of asymptotic statistical properties in terms of confidence intervals and likelihood ratio tests (LRTs) has been doubted (FREDKIN and RICE, 1992; ROSALES, 1999).

The maximum likelihood estimator (MLE) for parameter estimation in HMM has only recently been proven to be asymptotically normal under moderate conditions (BICKEL et al., 1998). The aim of the simulation study presented in this paper is to investigate finite sample properties of the MLE. Following a short overview of HMM theory (Section 2), we examine bias and variance of the estimator and confidence regions (Section 4; simulation settings in Section 3). We also investigate the power of likelihood ratio tests in different cases (Section 5 and Appendix) and apply parameter estimation as well as model selection by LRTs to measured biological data from sodium channels (Section 6).

2. Theory of Hidden Markov Models

2.1 *Parameterized HMM*

A parameterized discrete-time hidden Markov model $\mathcal{M}(\boldsymbol{\vartheta}) = (\boldsymbol{A}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\sigma})$ is related to two stochastic processes $\boldsymbol{X} = (X_t)_{1 \le t \le N}$ and $\boldsymbol{Y} = (Y_t)_{1 \le t \le N}$: The background process $\boldsymbol{X}$ of transitions between the m background states is a homogeneous Markov process which is governed by the $m \times m$ stochastic matrix $\boldsymbol{A}$ of transition probabilities $a_{ij} = P(X_t = j \mid X_{t-1} = i)$. Its initial distribution is given by $\pi_i = P(X_1 = i)$, the stationary distribution will be denoted by (p_i). The real-valued output $\boldsymbol{Y}$ follows $Y_t = \mu(X_t) + \sigma(X_t)\,\epsilon_t$ with state-dependent expectation value and standard deviation, and i.i.d. random variables $\epsilon_t \sim \mathcal{N}(0,1)$. The parameter vector is denoted by $\boldsymbol{\vartheta} \in \Theta \subseteq \mathbb{R}^n$. The log-likelihood $\mathcal{L}_{\mathcal{M}(\boldsymbol{\vartheta})}(\boldsymbol{Y})$ can be calculated efficiently by the forward-algorithm (BAUM et al., 1970; RABINER, 1989).

2.2 *Asymptotic properties of the MLE*

Under mild regularity conditions, the MLE $\hat{\boldsymbol{\vartheta}}$ for the true parameter vector $\boldsymbol{\vartheta}^0$ is asymptotically unbiased, normally distributed and consistent (LEROUX, 1992; BICKEL et al., 1998):

$$\sqrt{N}\,(\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^0) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{for} \quad \mathrm{N} \to \infty\,, \tag{1}$$

with

$$\boldsymbol{\Sigma}^{-1} = \lim_{N \to \infty} -\frac{1}{N} \left[\frac{\partial^2}{\partial \vartheta_i\, \partial \vartheta_j}\, \mathcal{L}_{\mathcal{M}(\boldsymbol{\vartheta})}(\boldsymbol{Y}) \bigg|_{\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0} \right]_{ij}.$$

Thus, $\dfrac{\partial^2}{\partial \vartheta_i\, \partial \vartheta_j}\, \mathcal{L}_{\mathcal{M}(\boldsymbol{\vartheta})}(\boldsymbol{Y}) \bigg|_{\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}}$ is an estimate of the inverse covariance matrix of $\hat{\boldsymbol{\vartheta}}$. The asymptotic normality yields the following asymptotic distribution for the LRT

test statistics (GIUDICI et al., 2000): Let $\mathcal{M}_1, \mathcal{M}_2$ be two nested models with parameter spaces $\Theta_1 \subseteq \mathbb{R}^{n_1}, \Theta_2 \subseteq \mathbb{R}^{n_2}, \Theta_1 \subset \Theta_2$. Then under the null hypothesis $\boldsymbol{\vartheta}^0 \in \Theta_1$, we have

$$2\big(\mathcal{L}_{\mathcal{M}_2}(\boldsymbol{Y}) - \mathcal{L}_{\mathcal{M}_1}(\boldsymbol{Y})\big) \sim \chi^2_{n_2-n_1}\,.$$

Identifiability of parameters as they occur for HMMs which model aggregated processes as well as problems due to boundaries of the parameter space have been studied elsewhere (KIENKER, 1989; WAGNER et al., 1999).

3. General Simulation Settings and Notations

We will use the following conventions for HMMs: Parameters for $\boldsymbol{A}$ are either the off-diagonal elements themselves or, in general, the entries of $\boldsymbol{A}$ are functions of the parameters. The diagonal elements are obtained by row normalization to unity. The output levels $\boldsymbol{\mu}$ are parameterized by a parameter μ_1 defining the lowest level and by offset-parameters $\Delta_2, \ldots, \Delta_m$ defining the differences to the remaining levels.

For all simulations in this paper, we used the pseudo random number generator ran2 (PRESS et al., 1992). The parameter estimation was performed by some initial steps of the EM algorithm, followed by a numerical optimization using NAG routine e04ucf (The Numerical Algorithms Group Ltd, Oxford, GB). The EM-algorithm for improved parameter re-estimation in parameterized HMM is given in MICHALEK and TIMMER (1999). During estimation, the general symmetry of changing the state numbering is broken by forcing the levels to stay in order.

The simulation parameters are denoted by ϑ_i^0, the ML-estimates of the ν-th simulation replication by $\hat{\vartheta}_i^\nu$ and their estimated standard deviation by $\hat{s}_i^\nu$. When model parameters are kept fixed during the estimation, this is marked by underlining the respective parameter in the description of the simulation setting, e.g. $\underline{\pi_1 = 1}$.

4. Consistency of the MLE

In this section, we investigate the finite sample properties of the MLE. We regard the bias b, the square root r of the mean square error (mse) and the coverage c of the 95%-confidence interval based on the estimated standard deviation of the estimator.

4.1 *Two-state model*

In our simulation study, we refer to the basic two-state model using the following generating parameters $\boldsymbol{\vartheta}^0$: $\underline{\pi_1 = 1}, \mu_1 = 0, \Delta_2 = 1$. The transition probabilities

Table 1

Simulation results for the MLE: bias b, square root r of mean square error and coverage c for the 95%-confidence region for $\hat{a}_{12}$ and $\hat{\Delta}$, respectively. Various parameters were used, with 2500 replications for each setting. Part I: data length $N = 100$ to $N = 3000$

σ	a_{12}	a_{21}	p_1	$\hat{a}_{12}$						$\hat{\Delta}$					
				$N = 100$ $b/r/c$	$N = 200$ $b/r/c$	$N = 400$ $b/r/c$	$N = 800$ $b/r/c$	$N = 1500$ $b/r/c$	$N = 3000$ $b/r/c$	$N = 100$ $b/r/c$	$N = 200$ $b/r/c$	$N = 400$ $b/r/c$	$N = 800$ $b/r/c$	$N = 1500$ $b/r/c$	$N = 3000$ $b/r/c$
0.4	0.10	0.10	0.500	0.0154	0.0054	0.0037	0.0014	0.0012	−0.00003	0.0006	0.0021	0.0007	0.0006	0.0001	0.0002
				0.0605	0.0378	0.0266	0.0178	0.0133	0.00889	0.0917	0.0630	0.0434	0.0304	0.0218	0.0154
				0.929	0.932	0.940	0.941	0.936	0.946	0.941	0.942	0.947	0.952	0.952	0.948
0.4	0.10	0.05	0.333	0.0273	0.0108	0.0047	0.0018	0.0008	0.0009	0.0050	0.0004	0.0007	−0.0000	0.0001	0.0000
				0.0861	0.0487	0.0315	0.0213	0.0152	0.0108	0.0960	0.0676	0.0464	0.0325	0.0232	0.0163
				0.916	0.944	0.938	0.942	0.947	0.947	0.942	0.946	0.949	0.946	0.953	0.950
0.4	0.10	0.01	0.091	0.059	0.052	0.0311	0.016	0.0084	0.0040	−0.004	−0.005	0.0007	0.0008	−0.0002	−0.0005
				0.135	0.130	0.0918	0.055	0.0337	0.0214	0.157	0.127	0.0858	0.0562	0.0395	0.0275
				0.926	0.928	0.938	0.943	0.951	0.950	0.932	0.938	0.951	0.950	0.958	0.955
0.4	0.05	0.05	0.500	0.0186	0.0068	0.0025	0.0017	0.00061	0.00040	0.002	0.00325	0.0018	−0.0004	0.0001	0.0004
				0.0598	0.0297	0.0182	0.0123	0.00887	0.00618	0.110	0.0626	0.0432	0.0300	0.0215	0.0156
				0.913	0.921	0.945	0.952	0.944	0.954	0.940	0.949	0.946	0.944	0.944	0.942
0.4	0.05	0.01	0.167	0.049	0.0307	0.0162	0.0077	0.0041	0.0020	−0.005	0.0001	0.0009	0.0007	0.0002	0.0002
				0.117	0.0817	0.0491	0.0260	0.0171	0.0109	0.133	0.0941	0.0628	0.0418	0.0286	0.0203
				0.929	0.925	0.943	0.955	0.950	0.945	0.935	0.937	0.941	0.941	0.957	0.950
0.4	0.01	0.01	0.500	0.059	0.0185	0.0056	0.00158	0.00068	0.00036	−0.194	−0.068	−0.0107	0.0002	0.0004	0.0002
				0.132	0.0616	0.0251	0.00652	0.00417	0.00280	0.316	0.204	0.0940	0.0308	0.0209	0.0146
				0.804	0.933	0.908	0.929	0.934	0.944	0.808	0.907	0.945	0.950	0.953	0.950
0.4	0.05	0.05	0.500	0.0186	0.0068	0.0025	0.0017	0.00061	0.00040	0.002	0.0033	0.0018	−0.0004	0.0001	0.0004
				0.0598	0.0297	0.0182	0.0123	0.00887	0.00618	0.110	0.0626	0.0432	0.0300	0.0215	0.0156
				0.913	0.921	0.945	0.952	0.944	0.954	0.940	0.949	0.946	0.944	0.944	0.942
1.0	0.05	0.05	0.500	0.052	0.0240	0.0113	0.0047	0.0022	0.00074	0.119	0.068	0.028	0.012	0.0078	0.0045
				0.114	0.0739	0.0392	0.0221	0.0146	0.00974	0.373	0.237	0.144	0.094	0.0669	0.0461
				0.892	0.875	0.916	0.934	0.944	0.947	0.873	0.915	0.929	0.944	0.947	0.951
2.0	0.05	0.05	0.500	0.103	0.069	0.052	0.0268	0.0150	0.0050	0.39	0.415	1.22	0.230	0.133	0.047
				0.181	0.159	0.127	0.0938	0.0630	0.0332	1.20	0.911	1.90	0.437	0.301	0.161
				0.872	0.800	0.789	0.803	0.842	0.884	0.752	0.809	0.741	0.882	0.913	0.934
4.0	0.05	0.05	0.500	0.136	0.106	0.096	0.074	0.066	0.043	0.54	0.93	1.22	1.19	1.03	0.77
				0.207	0.197	0.187	0.169	0.167	0.135	2.95	2.35	1.90	1.62	1.43	1.14
				0.883	0.801	0.751	0.702	0.678	0.679	0.734	0.739	0.741	0.759	0.800	0.835

Table 2
Simulation results for the MLE (see Table 1). Part II: data length $N = 10^4$ to $N = 3 \cdot 10^5$

σ	a_{12}	a_{21}	p_1	$\hat{a}_{12}$				$\hat{\Delta}$			
				$N = 10000$ $b/r/c$	$N = 30000$ $b/r/c$	$N = 100000$ $b/r/c$	$N = 300000$ $b/r/c$	$N = 10000$ $b/r/c$	$N = 30000$ $b/r/c$	$N = 100000$ $b/r/c$	$N = 300000$ $b/r/c$
1.0	0.05	0.05	0.500	0.00018	0.00014	0.000005	0.000008	0.0007	0.0003	0.00001	0.00010
				0.00528	0.00300	0.00161	0.000964	0.0257	0.0148	0.00819	0.00458
				0.945	0.949	0.958	0.947	0.951	0.953	0.946	0.945
2.0	0.05	0.05	0.500	0.0013	0.00028	0.000138	0.00002	0.0118	0.0020	0.0012	0.0004
				0.0134	0.00738	0.00403	0.00226	0.0773	0.0429	0.0231	0.0135
				0.935	0.946	0.943	0.949	0.942	0.950	0.954	0.948
4.0	0.05	0.05	0.500	0.0210	0.0060	0.00114	0.00026	0.367	0.114	0.0291	0.0074
				0.0897	0.0415	0.0161	0.00894	0.646	0.237	0.0884	0.0480
				0.764	0.859	0.917	0.935	0.891	0.944	0.952	0.956
8.0	0.05	0.05	0.500	0.079	0.047	0.026	0.0073	2.15	1.65	0.99	0.389
				0.187	0.158	0.111	0.0577	2.69	2.33	1.52	0.598
				0.600	0.582	0.677	0.762	0.766	0.811	0.881	0.946

a_{12}, a_{21} as well as the noise level parameter $\sigma = \sigma_1 = \sigma_2$ are chosen in the following range: $a_{12}, a_{21} = 0.01, 0.05, 0.1$; $\sigma = 0.4, 1.0, 2.0, 4.0, 8.0$, see Tables 1 and 2. Note that with $a_{12} = a_{21}$ the marginal distribution of Y_t becomes unimodal for $\sigma \geq 0.5$. For $n = 2500$ replications for each setting, we estimated the bias $\hat{b} := \frac{1}{n} \sum_{\nu=1}^{n} (\hat{\vartheta}_i^{\nu} - \vartheta_i^0)$, the mse $\hat{r}^2 := \frac{1}{n} \sum_{\nu=1}^{n} (\hat{\vartheta}_i^{\nu} - \vartheta_i^0)^2$ and the coverage $\hat{c} := \frac{1}{n} \sum_{\nu=1}^{n} 1\{\vartheta_i^0 \in [\hat{\vartheta}_i^{\nu} - 1.96\,\hat{\sigma}_i^{\nu}, \hat{\vartheta}_i^{\nu} + 1.96\,\hat{\sigma}_i^{\nu}]\}$ (where $1\{\cdot\}$ is an indicator function) related to both $\vartheta \equiv a_{12}$ and $\vartheta \equiv \Delta$. For $n = 2500$, their respective standard errors, se $(.)$, are se $(\hat{r}) \approx 0.014\,\hat{r}$, se $(\hat{b}) \approx 0.02\,\mathrm{se}(\hat{r})$, and, for $\hat{c} = 0.95$, se $(\hat{c}) \approx 0.004$.

We observe a general tendency of overestimating a_{12}, i.e. the estimates tend to assume more transitions than actually happened. If a_{12} is small or σ is large, the relative bias can exceed 100%, even for sample sizes of several thousand data points. Estimates of Δ are also positively biased. Only if a_{12} is very small, the opposite tendency can be observed for small samples. The bias can exceed 100%

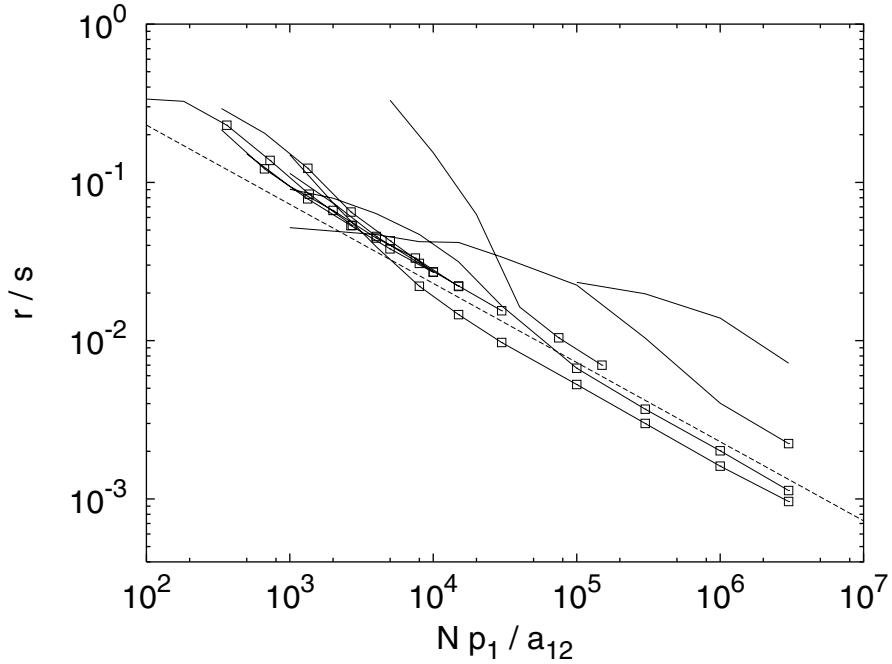

Fig. 1. The scaling behavior of the square root r of the mean square error, with N the length of the time series, σ the noise standard deviation, a_{12} the transition probability, and p_1 the stationary probability of $X = 1$. All values from Tables 1 and 2 are plotted, with lines connecting series of increasing N. Only those points are marked, where a coverage of $c > 0.933$ indicates that an asymptotical behavior already is given. The dashed line is $2.3\sqrt{a_{12}/(Np_1)}$

for large values of σ. For $\sigma = 0.4$, the coverage frequency of the confidence intervals is close to the nominal level even for small sample sizes; however, we can observe an anti-conservative tendency. For $\sigma \geq 1$, large sample sizes are necessary for a sufficient approximation. This holds for both a_{12} and Δ.

We also investigated HMM with $m = 4$ states and $\underline{\pi_1 = 1}$, $\mu_1 = 1$, $\Delta_2 = 1$, $\Delta_3 = 2$, $\Delta_4 = 3$. The results were comparable to the two-state case.

4.2 *Scaling of the estimation errors*

The rate of convergence of the MLE is asymptotically given by equation (1). However, the finite-sample scaling behavior as well as the absolute estimation variance depend on the true parameters. In order to give a helpful a-priori appraisal of the estimation accuracy which one can expect for a time series of certain length, we investigate the scaling behavior of the mse. Figure 1 shows r/σ plotted vs. Np_1/a_{12}.

Thus, over a wide range of parameters, the expected root mean square error of $\hat{a}_{12}$ is proportional to $\sigma\sqrt{a_{12}}/\sqrt{Np_1}$: the mse decreases linearly with the total number of time points where the process occupies the respective state, Np_1, multiplied by $1/a_{12}$, which is the mean occupancy time τ_i. It also increases linearly with the noise variance. For the constellations considered, the proportionality factor was between 1.5 and 4.

5. Applicability of the LRT

In the following simulations, we examine the properties of the LRT. We investigate (a) the percentage of rejection under the null hypothesis, and (b) the power against violations of the null hypothesis. The nominal significance level of the applied tests was always $\alpha = 0.05$. The settings of the test situations are motivated by ion channel analysis, see e.g. KLEIN et al. (1997); TIMMER and KLEIN (1997).

5.1 Testing for equality of transition probabilities

The model parameters for simulating were: $m = 3$, $\underline{\pi_1 = 1}$, $\underline{\mu_1 = 1}$, $\underline{\Delta_2 = 1}$, $\underline{\Delta_3 = 2}$, $\underline{\sigma^2 = 1}$. The transition probabilities were $a_{ij} = 0.05$ or $\underline{a_{ij} = 0.025}$. The two fitted models assumed all a_{ij} to be equal (null hypothesis), and arbitrary entries for a_{ij} (alternative hypothesis), respectively. Thus, the rejection range for the twofold log-likelihood difference starts at $\chi^2_{5;\,0.05} = 11.07$. Table 3 gives the frequency of rejections for $n = 2500$ replications for different σ, a_{ij} and N; the accuracy is 0.004 for a rejection frequency of 0.05. Actual and nominal level are close

Table 3
Estimated actual level of the LRT for equality of all transition probabilities, using different parameters for the noise level σ, the mean occupancy time $1/a_{ii}$ and different data length N. The nominal level was 0.05

σ	a_{ii}	$N = 1000$	$N = 5000$	$N = 10000$	$N = 50000$
0.4	0.9	0.052	0.050	0.050	0.042
	0.95	0.061	0.046	0.050	0.050
1.0	0.9	0.053	0.048	0.051	0.048
	0.95	0.066	0.057	0.055	0.053
2.0	0.9	0.016	0.042	0.060	0.055
	0.95	0.024	0.049	0.050	0.056

together, at least for $N \geq 5000$. For large values of σ, we observe a conservative behavior in smaller samples.

The power of this LRT is investigated for the case $\sigma = 1.0, a_{ij} = 0.05$ with $N = 2500$ and $N = 10000$. The following violation of the null hypothesis is examined: The mean occupancy time τ_1 of state 1 is changed to $\tau_1' = \epsilon\tau_0$, while τ_3 is changed to $\tau_3' = \tau_0/\epsilon$ and $\tau_2 = \tau_0$ stays unchanged. The frequency of rejections of the null hypothesis is plotted in Figure 2 for different values of ϵ in the neighborhood of 1. We conclude that we have – for $N = 10000$ – sufficient power to detect even small deviations from the null hypothesis.

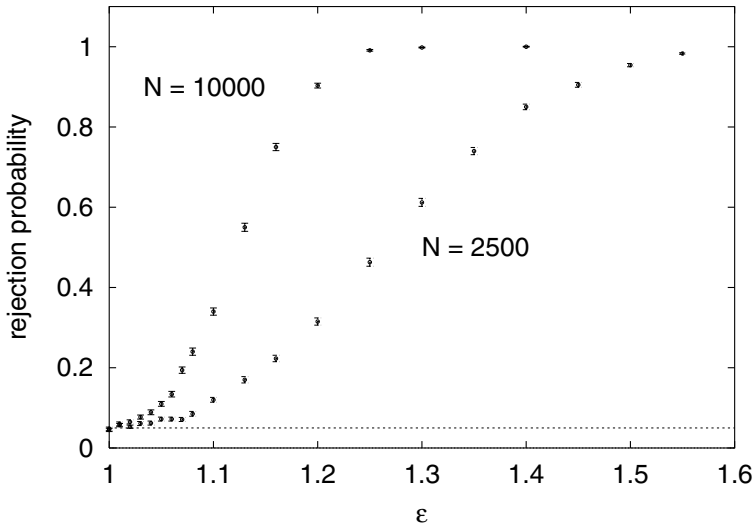

Fig. 2. Power of the LRT in the three-state HMM for the case of different mean occupancy times $\tau_1 = \epsilon\tau_0, \tau_2 = \tau_0, \tau_3 = \tau_0/\epsilon$ for two different data lengths. The dashed line marks the nominal test size of 5%

5.2 *Testing in the case of superimposed Markov processes*

In this section, we examine the superposition of two HMMs with each $m = 2$ states. We denote their parameters by superscripts i and ii, respectively. In the sequel we also write $a^{i/ii} := a_{12}^{i/ii}, b^{i/ii} := a_{21}^{i/ii}$, and $\bar{\alpha} := 1 - \alpha$ for probabilities α. As their independent superposition, we obtain a new HMM $\tilde{\mathcal{M}}$ with $\tilde{m} = 4$, $\tilde{\mu}_1 = \mu_1^i + \mu_1^{ii}$, $\tilde{\Delta}_2 = \Delta_2^i$, $\tilde{\Delta}_3 = \Delta_2^{ii}$, $\tilde{\Delta}_4 = \Delta_2^i + \Delta_2^{ii}$, $\tilde{\sigma}^2 = \sigma^{i^2} + \sigma^{ii^2}$, and transition matrix $\tilde{\boldsymbol{A}}$. The superposition model is shown in Table 4 together with $\tilde{\boldsymbol{A}}$. Under the null hypothesis for the following tests the two processes have the same parameters. In this case, $\tilde{\mathcal{M}}$ can be replaced by an equivalent model $\tilde{\mathcal{M}}^0$ with $m^0 = 3$, $\tilde{\mu}_1^0 = 2\mu_1^i$, $\tilde{\Delta}_2^0 = \Delta_2^i$, $\tilde{\Delta}_3^0 = 2\Delta_2^i$, $\tilde{\sigma}^0 = \sqrt{2}\,\sigma^i$ that produces the same observable process when $\tilde{\boldsymbol{A}}$ is replaced by $\tilde{\boldsymbol{A}}^0$ (see Table 4).

For the simulations, we choose under H_0: $\underline{\pi_1 = 1}$, $\underline{\mu_1 = 1}$, $a = 0.05$, $b = 0.09$, $\Delta_2 = 1.0$, $\underline{\tilde{\sigma} = 1.0}$, and $N = 10000$. Under the alternative, the parameters a, b, Δ_2 of the two models may be different.

5.2.1 *Testing for identical parameters*

We first consider the test of $\tilde{\mathcal{M}}$ (possibly differing parameters) vs. $\tilde{\mathcal{M}}^0$ (identical parameters). The rejection range for the LRT starts at $\chi^2_{3;\,0.05} = 7.8147$. The first violation of the null hypothesis regarded assumes one Markov process to exhibit shorter dwell times than the other. With $a^i = \epsilon a, b^i = \epsilon b, a^{ii} = a/\epsilon, b^{ii} = b/\epsilon$, the stationary distributions are unchanged, and so is the stationary distribution of the superposition. Also, the mean occupancy times of the three resulting states are unchanged (see Appendix). The LRT is less sensitive to such a change in the dynamics (Figure 3).

Another violation of the null hypothesis assumes the two Markov processes possess different output levels. With $\Delta_2^i = \epsilon\Delta_2, \Delta_2^i = \Delta_2/\epsilon$, the intermediate level

Table 4

The superposition of two hidden Markov models: The resulting output levels dependent on the background states X_t^i, X_t^{ii} of the two processes, and the resulting transition matrix $\tilde{\boldsymbol{A}}$. The transition matrix is also given for the special case of identical parameters which is the null hypothesis for the investigated tests ($\tilde{\boldsymbol{A}}^0$). 'x' denotes the diagonal entries obtained by normalization.

X_t^i, X_t^{ii}	level
2, 2	$-\mu_1^i + \mu_1^{ii} + \Delta_2^i + \Delta_2^{ii}$
2, 1	$-\mu_1^i + \mu_1^{ii} + \Delta_2^{ii}$
1, 2	$-\mu_1^i + \mu_1^{ii} + \Delta_2^i$
1, 1	$-\mu_1^i + \mu_1^{ii}$

$$\tilde{\boldsymbol{A}} = \begin{pmatrix} \mathrm{x} & a^i\bar{a}^{ii} & \bar{a}^i a^{ii} & a^i a^{ii} \\ b^i\bar{a}^{ii} & \mathrm{x} & b^i a^{ii} & \bar{b}^i a^{ii} \\ \bar{a}^i b^{ii} & a^i b^{ii} & \mathrm{x} & a^i\bar{b}^{ii} \\ b^i b^{ii} & \bar{b}^i b^{ii} & b^i\bar{b}^{ii} & \mathrm{x} \end{pmatrix}, \quad \tilde{\boldsymbol{A}}^0 = \begin{pmatrix} \mathrm{x} & 2a\bar{a} & a^2 \\ \bar{a}b & \mathrm{x} & a\bar{b} \\ b^2 & 2b\bar{b} & \mathrm{x} \end{pmatrix}$$

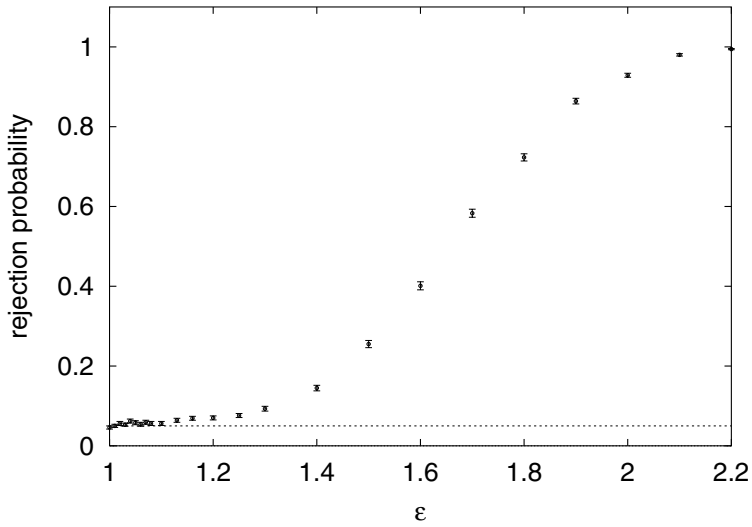

Fig. 3. Power of the LRT for two superimposed HMMs for the case of different time constants $\tau^i_{1/2} = \tau^0_{1/2}/\epsilon$, $\tau^{ii}_{1/2} = \epsilon\tau^0_{1/2}$, and $N = 10\,000$. The dashed line marks the nominal test size of 5%

splits up into two separate levels. The LRT is rather sensitive to such a change in the stationary output distribution (Figure 4).

In both cases, actual and nominal level are close together under H_0.

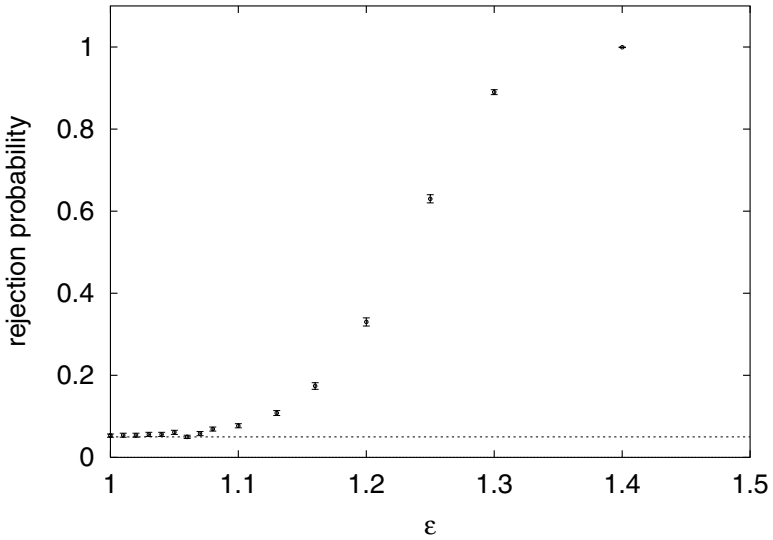

Fig. 4. Power of the LRT for two superimposed HMMs for the case of different output levels $\Delta^i_1 = \epsilon\Delta^0_1$, $\Delta^{ii}_1 = \Delta^0_1/\epsilon$, and $N = 10\,000$. The dashed line marks the nominal test size of 5%

5.2.2 *Testing for independence of superpositions*

The following violation of the above null hypothesis allows for dependent dynamics of the two Markov process, assuming that μ and σ still coincide: The transition probability a is changed to $a' = \epsilon_a a$ if the other process is in its state 2 (this is relevant for the states 1,2 and 2,1 of the superposition), and also b is changed to $b' = \epsilon_b b$ if the other process is in its state 2 (relevant for state 2,2 of the superposition). We investigate the power of the corresponding LRT by varying the single parameter $\epsilon = \epsilon_a = 1/\epsilon_b$ in the neighborhood of 1. We use for the superposition model $\tilde{\mathcal{M}}^{\text{alt}}$ with $m = 3, \underline{\mu_1 = 1}, \underline{\Delta_2 = 1}, \underline{\Delta_3 = 2}, \underline{\sigma = 1}$. The transition matrix for the simulation as well as for the alternative hypothesis is:

$$\tilde{\boldsymbol{A}}^{\text{alt}} = \begin{pmatrix} X & 2a\bar{a} & a^2 \\ \bar{a}'b & X & a'\bar{b} \\ b'^2 & 2b'\bar{b}' & X \end{pmatrix}.$$

For estimation under the alternative, a, a', b, b' are free parameters, thus the quantile for testing $\tilde{\mathcal{M}}^{\text{alt}}$ vs. $\tilde{\mathcal{M}}^0$ is $\chi^2_{2;\,0.95} = 5.991$. Under this alternative, the mean occupancy times as well as the stationary distribution are changed compared with the null hypothesis. As we can conclude from Figure 5, the LRT is rather sensitive.

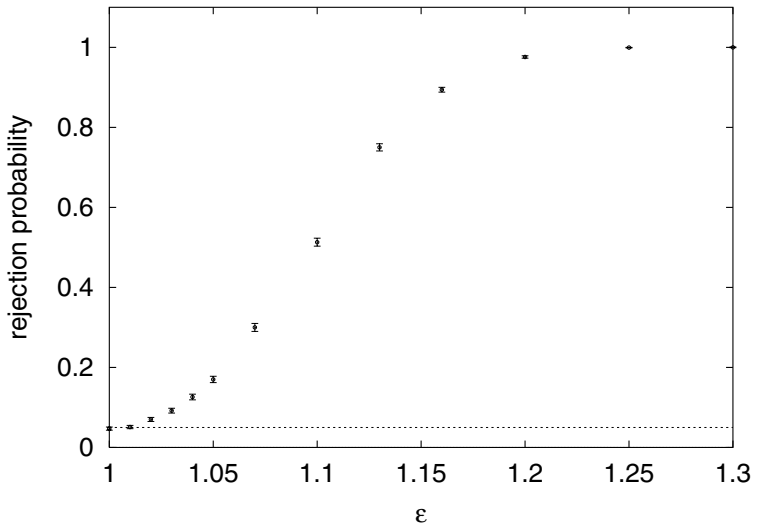

Fig. 5. Power of the LRT for two superimposed HMMs for the case of dependent dynamics: $a'_{12} = \epsilon a_{12}$ and $b'_{21} = b_{12}/\epsilon$ if the other process is in its state 2 ($N = 10\,000$). The dashed line marks the nominal test size of 5%

6. Application to Measured Time Series from Sodium Channels

Recordings of single human muscle Na^+-channels were analyzed using a generalized HMM incorporating filtered noise. The F1473S mutant was expressed stably in a cell line (FLEISCHHAUER et al., 1998). The so-called cell-attached patches were measured using a voltage protocol of hyper-polarization changing to depolarization, which was applied for several hundred repetitions. The generalized model was necessary in order to handle the high level of observational noise together with the filtering of this physiological measurements (MICHALEK et al., 2000). Some traces of the data are shown in Figure 6, the recording procedure is described in MICHALEK et al. (1999). The physiological interest in the behavior of single ion channels arises from their fundamental role in the cell's interaction with its environment as well as from their responsibility for certain muscle diseases (HOFFMANN et al., 1995).

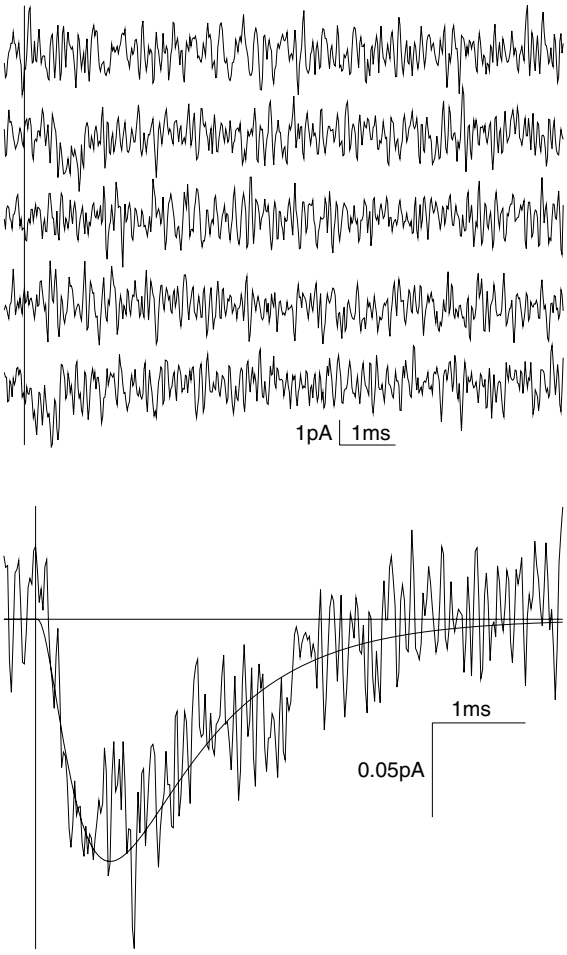

Fig. 6. (above) Representative raw current traces of single channel recordings. Openings (plotted downwards) occur within the first two milliseconds of traces 2 and 5. (below) Theoretical time course of the average current from the fitted generalized HMM (smooth) for comparison with the average over the 600 recorded traces (vertical line at start of depolarization). The theoretical curve is scaled for equal peak height

The main characteristics of ion channel behavior can be described by Markov processes of transitions between classes of molecular conformations, called states. Thus, the interest is to identify the so-called gating-scheme, which governs the connectivity of the Markov states together with the respective rate constants of the transitions (MICHALEK and TIMMER, 1999). This task was performed by subsequent LRTs, here.

A simple model that has often been used in the past to describe the channel gating contains one open, one inactivated and several closed states (HORN and LANGE, 1983; HORN and VANDENBERG, 1984; PATLAK, 1991; CHAHINE et al., 1994), see model number 2 in Table 5: Due to the experimental preparation, the channels are forced to start from the outmost left closed state at each repetition of the voltage protocol. The so-called inactivated states are closed as well. They are nearly absorbing in terms of the Markov dynamics. However, the number of closed states and the number of different open states which contribute to the observed dynamics is not known beforehand. We subsequently tested against simpler and more complex models by means of LRTs. The resulting log-likelihood values

Table 5
Log-likelihood values for different gating schemes as described in the text.

gating scheme		log-likelihood
no.	transition graph	
1	$C \leftrightarrows O$, $C \leftrightarrows I$	−52383.45
2	$C_1 \leftrightarrows C_2 \leftrightarrows O$, $C_2 \leftrightarrows I$	−52366.32
3	$C_1 \to C_2 \leftrightarrows O$, $C_2 \leftrightarrows I$	−52367.04
4	$C_1 \to C_2 \to C_3 \leftrightarrows O$, $C_3 \leftrightarrows I$	−52367.99
5	$O_1 \leftrightarrows C_1 \to C_2 \leftrightarrows O$, $C_2 \leftrightarrows O_2$	−52364.33

are given in Table 5 and the resulting decisions are:

$$1 \rightarrow 2 \rightarrow \boxed{3} \begin{array}{c} 4 \\ \downarrow \\ \\ \uparrow \\ 5 \end{array}.$$

Each arrow indicates a LRT with a significant result for the more complex model or, otherwise, the case that the null hypothesis could not be rejected and that the less complex model is sufficient for the statistical description of the data. Details of the analysis, tests of the method's consistency and physiological interpretations are described in MICHALEK et al. (1999). The best model number 3 was able to reproduce the current time course of the data, see Figure 6.

7. Conclusions

We investigated the finite sample size properties of maximum likelihood estimators and likelihood ratio tests for hidden Markov models: (a) For a wide range of parameter values, we give an approximative scaling behavior for the accuracy of parameter estimates. (b) The results of several likelihood ratio tests do allow for a comparison of the test's power under different violations of the null hypothesis. The likelihood ratio tests are more sensitive for changes in the stationary output distribution as well as in the mean occupancy times of the levels. It is less sensitive for higher-order changes, i.e. if only the correlations between pairs of adjacent occupancy times are affected.

We successfully applied our results to data measured from sodium channels of human skeletal muscle. With those criteria concerning the applicability of likelihood methods, hidden Markov models together with proposed extensions (HAMILTON, 1994; ALBERTSEN and HANSEN, 1994; KIM, 1994; VENKATARAMANAN et al., 1998; MICHALEK et al., 1999, 2000) allow for investigating dynamical systems that cannot be observed directly in a statistically reliable way.

Appendix

The ϵ-parameterized matrix $\tilde{\boldsymbol{A}}'(\epsilon)$ used for the simulations in Section 5.2.1 has the following derivative with respect to ϵ:

$$\frac{\partial}{\partial \epsilon} \tilde{\boldsymbol{A}}'(\epsilon) = \begin{pmatrix} -a^i + a^{ii}/\epsilon^2 & a^i & -a^{ii}/\epsilon^2 & 0 \\ b^i & -b^i + a^{ii}/\epsilon^2 & 0 & -a^{ii}/\epsilon^2 \\ -b^{ii}/\epsilon^2 & 0 & b^{ii}/\epsilon^2 - a^i & a^i \\ 0 & -b^{ii}/\epsilon^2 & b^i & b^{ii}/\epsilon^2 - b^i \end{pmatrix},$$

and for $\epsilon \approx 1$ (and therefore $a^i = a^{ii} = a$, $b^i = b^{ii} = b$):

$$\approx \begin{pmatrix} 0 & a & -a & 0 \\ b & -b+a & 0 & -a \\ -b & 0 & b-a & a \\ 0 & -b & b & 0 \end{pmatrix}. \tag{19}$$

Thus, the only effects for small ϵ are (a) that the occupancy time distribution for the intermediate level 2 splits up and is now governed by two different time constants (for $a \neq b$) while the mean occupancy time is nearly unchanged, and (b) that the probability for subsequent transitions from the lowest level 1 to level 2 and back increases compared to $1 \to 2 \to 3$. The same is true for level 3 (highest) $\to 2 \to 3$ compared to $3 \to 2 \to 1$.

The log-likelihood is less sensitive to such changes compared e.g. to changes in mean occupancy times or in the probability for direct transitions. Thus, the power of the LRT is poor, here.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG Ho 496/4-2).

References

ALBERTSEN, A. and HANSEN, U.-P., 1994: Estimation of kinetic rate constants from multi-channel recordings by a direct fit of the time series. *Biophys. J.* **67**, 1393–1403.

BAUM, L. E., PETRIE, T., SOULES, G., and WEISS, N., 1970: A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.* **41**, 164–171.

BECKER, J., HONERKAMP, J., HIRSCH, J., SCHLATTER, E., and GREGER, R., 1994: Analyzing ion channels with hidden Markov models. *Pfl. Arch.* **426**, 328–332.

BICKEL, P., RITOV, Y., and RYDÉN, T., 1998: Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Stat.* **26**, 1614–1635.

CHAHINE, M., GEORGE, A. L., ZHOU, M., JI, S., SUN, W., BARCHI, R. L., and HORN, R., 1994: Sodium channel mutations in paramyotonia congenita uncouple inactivation from activation. *Neuron* **12**, 281–294.

CHUNG, S.-H. and. KENNEDY, R. A, 1996: Coupled Markov chain model: Characterization of membrane channel currents with multiple conductance sublevels as partially coupled elementary pores. *Math. Biosc.* **133**, 111–137.

CHUNG, S.-H., MOORE, J., XIA, L., PREMKUMAR, L. S., and GAGE, P. W., 1990: Characterization of single channel currents using digital signal processing techniques based on hidden Markov models. *Phil. Trans. R. Soc. Lond. B* **329**, 265–285.

COX, D. and HINKLEY, C., 1974: *Theoretical Statistics*. Chapman & Hall.

DURBIN, R., EDDY, R., KROGH, A., and MITCHISON, G., 1998: *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge University Press.

ELLIOTT, R., AGGOUN, L., and MOORE, J., 1995: *Hidden Markov models: Estimation and control*. Springer.

FLEISCHHAUER, R., MITROVIĆ, N., DEYMEER, F., GEORGE, A. L., LEHMANN-HORN, F., and LERCHE, H., 1998: Effects of temperature and mexiletine on a sodium channel mutation causing paramyotonia congenita. *Pfl. Arch.* **436**, 757–765.

FREDKIN, D. R. and RICE, J. A., 1992: Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. R. Soc. Lond. B* **249**, 125–132.

GIUDICI, P., RYDÉN, T., and VANDEKERKHOVE, P., 2000: Likelihood-ratio tests for hidden Markov models. *Biometrics* **56**, 742–747.

HAMILTON, J., 1990: Analysis of time series subject to changes in regime. *J. Econometrics* **45**, 39–70.

HAMILTON, J., 1994: *Time Series Analysis*. Princeton University Press.

HOFFMANN, E. P., LEHMANN-HORN, F., and RÜDEL, R., 1995: Overexited or inactive: Ion channels in muscle disease. *Cell* **80**, 681–686.

HORN, R. and LANGE, K., 1983: Estimating kinetic constants from single channel data. *Biophys. J.* **43**, 207–223.

HORN, R. and VANDENBERG, C., 1984: Statistical properties of single sodium channels. *J. Gen. Physiol.* **84**, 505–534.

KIENKER, P., 1989: Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B* **236**, 269–309.

KIM, C.-J., 1994: Dynamic linear models with Markov-switching. *J. Econometrics* **60**, 1–22.

KLEIN, S., TIMMER, J., and HONERKAMP, J., 1997: Analysis of multichannel patch clamp recordings by hidden Markov models. *Biometrics* **53**, 870–884.

KROGH, A., BROWN, M., MIAN, I., SJÖLANDER, K., and HAUSSLER, D., 1994: Hidden Markov models in computational biology: Applications to protein modeling. *J. Mol. Biol.* **235**, 1501–1531.

LEROUX, B. G., 1992: Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.* **40**, 127–143.

MICHALEK, S. and TIMMER, J., 1999: Estimating rate constants in hidden Markov models by the EM algorithm. *IEEE Trans. Sig. Proc.* **47(1)**, 226–228.

MICHALEK, S., LERCHE, H., WAGNER, M., MITROVIĆ, N., SCHIEBE, M., LEHMANN-HORN, F., and TIMMER, J., 1999: On identification of Na^+ channel gating schemes using moving-average filtered hidden Markov models. *Eur. Biophys. J* **28**, 605–609.

MICHALEK, S., WAGNER, M., and TIMMER, J., 2000: A new approximate likelihood estimator for ARMA-filtered hidden Markov models. *IEEE Trans. Sig. Proc.* **48(6)**, 1537–1547.

MILBURN, T., SAINT, D. A., and CHUNG, S.-H., 1995: The temperature dependence of conductance of the sodium channel: implications for mechanisms of ion permeation. *Receptors Channels* **3(3)**, 201–211.

PATLAK, J. (1991): Molecular kinetics of voltage-dependent Na^+ channels. *Physiol. Rev.* **71**, 1047–80.

PRESS, W., FLANNERY, B., SAUL, B., and. WETTERLING, W, 1992: *Numerical Recipes*. Cambridge University Press.

RABINER, L. R., 1989: A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–285.

ROSALES, R., 1999: *Bayesian Analysis of Hidden Markov Models for Single Ion Channel Records*. Dissertation, University of Cambridge.

TIMMER, J. and KLEIN, S., 1997: Testing the Markov condition in ion channel recordings. *Phys. Rev.* E **55(3)**, 3306–3311.

VENKATARAMANAN, L., WALSH, J. L., KUC, R., and SIGWORTH, F., 1998: Identification of hidden Markov models for ion channel currents, part I: Colored, backgroand noise. *IEEE Trans. Sig. Proc.* **46(7)**, 1901–1915.

WAGNER, M., MICHALEK, S., and TIMMER, J., 1999: Estimating transition rates in aggregated Markov models of ion-channel gating with loops and with nearly equal dwell times. *Proc. R. Soc. Lond.* B **266**, 1919–1926.

Dr. JENS TIMMER
Center for Data Analysis
University of Freiburg
Eckerstraße 1
79 104 Freiburg
Germany
jeti@fdm.uni-freiburg.de
Phone: +49 761 203 5829
Fax: +49 761 203 7700

Received, December 1999
Revised, August 2000
Revised, January 2001
Revised, May 2001
Accepted, May 2001