ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2020)

Inference for biomedical data by using diffusion models with covariates and mixed effects

Mareile Große Ruse,

University of Copenhagen, Denmark

Adeline Samson

Université Grenoble Alpes, France

and Susanne Ditlevsen

University of Copenhagen, Denmark

[Received November 2017. Revised September 2019]

Summary. Neurobiological data such as electroencephalography measurements pose a statistical challenge due to low spatial resolution and poor signal-to-noise ratio, as well as large variability from subject to subject. We propose a new modelling framework for this type of data based on stochastic processes. Stochastic differential equations with mixed effects are a popular framework for modelling biomedical data, e.g. in pharmacological studies. Whereas the inherent stochasticity of diffusion models accounts for prevalent model uncertainty or misspecification, random-effects model intersubject variability. The two-layer stochasticity, however, renders parameter inference challenging. Estimates are based on the discretized continuous time likelihood and we investigate finite sample and discretization bias. In applications, the comparison of, for example, treatment effects is often of interest. We discuss hypothesis testing and evaluate by simulations. Finally, we apply the framework to a statistical investigation of electroencephalography recordings from epileptic patients. We close the paper by examining asymptotics (the number of subjects going to ∞) of maximum likelihood estimators in multi-dimensional, non-linear and non-homogeneous stochastic differential equations with random effects and included covariates.

Keywords: Approximate maximum likelihood; Asymptotic normality; Consistency; Covariates; Electroencephalography data; Local asymptotic normality; Mixed effects; Non-homogeneous observations; Random effects; Stochastic differential equations

1. Introduction

Many biomedical studies are based on image data, which are characterized by a high time resolution, but also a low signal-to-noise ratio. The same happens with electroencephalography data, which are measurements of electrical activitity measured from electrodes on the scalp and are proxies of underlying brain activity. This high frequency and noisy nature of the data lends itself naturally to be modelled by continuous time stochastic processes. Moreover, data are often multi-dimensional and repeated on a collection of subjects. The noise may be due to factors such as internal and external fluctuations, difficult experimental conditions or a collection of multiple unmeasured effects, e.g. non-specified feedback mechanisms or genetic

Address for correspondence: Susanne Ditlevsen, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark.
E-mail: susanne@math.ku.dk

 0035–9254/20/69000

variation. The intrasubject variability in longitudinal data asks for a model that incorporates system noise. Any systematic intersubject variability is usually well explained by the inclusion of covariate information, e.g. treatment regime, gender or specifics of experimental conditions. The remaining intersubject variability can then be taken care of by random effects. The main goal is to model electroencephalography measurements from multiple channels and to draw statistical inference about the underlying dynamics, but we are also motivated by a compartment model arising in a recent pharmacological study based on image data. Both types of data are measured at high frequency, i.e. the sampling frequency is fast compared with the typical timescales of the observed system. This enables us to employ techniques facilitating the use of continuous time stochastic processes. We therefore propose a new modelling framework where the observed time series are assumed to be generated from a multi-dimensional stochastic differential equation (SDE), which accounts for systematic and random intersubject variability through covariates and random effects.

The brain consists of a complex network of interconnected regions, with interactions that evolve dynamically during normal activity, and which shows pathological patterns during epileptic seizures. A common tool to infer the dynamical patterns is multisensor recordings of brain activity, such as electroencephalography recordings. They are often used to diagnose epilepsy, which causes abnormalities in the electroencephalography measurements. It is of great interest to establish the functional network giving rise to the measured electroencephalography data, and most methods are non-parametric using correlations and signal processing techniques (Prado *et al.*, 2001; Qin and Wang, 2008; Stephen *et al.*, 2014; Zhang *et al.*, 2015; Wulsin *et al.*, 2016; Ruegamer *et al.*, 2018). Here, we take a different road, modelling parametrically the data by interacting stochastic processes, where the estimated parameters yield measures of the connectivity and the changes that are incurred during epileptic seizures. This can serve as alternative and supplementary measures to characterize the dynamics. Furthermore, the modelling of electroencephalography data, and in particular epileptic seizures, is challenging because of the large amount of heterogeneity in the electroencephalography signal between channels and across individuals. Thus, the standard way is to do individual analyses on each subject. We shall instead approach this with random effects to gain statistical power and to find common characteristics over the population.

Models that combine SDEs and random effects (i.e. so-called stochastic differential mixed effects models (SDMEMs)) have become a popular framework for modelling biological data (Guy *et al.*, 2015; Donnet *et al.*, 2010; Møller *et al.*, 2010; Leander *et al.*, 2014; Picchini *et al.*, 2008; Picchini and Forman, 2019). They come with three advantages: firstly, they capture intersubject variations by incorporation of random effects. Secondly, they account for model uncertainty or environmental fluctuations by their inherent stochasticity. Lastly, they remedy the otherwise omnipresent issue of the inconsistent drift estimator (Kessler *et al.*, 2012) in plain SDEs (only fixed effects), when the observation time horizon is finite, because the mixed effects approach facilitates pooling of data across subjects, which leads to unbiasedness of the drift estimator as the number of subjects approaches ∞.

However, the flexibility and robustness of SDMEMs come at a price and bear particular challenges in terms of statistical inference. The data likelihood in these models is generally intractable, for two reasons: on the one hand, the likelihood for (non-linear) SDE models is analytically not available, rendering parameter inference for standard SDE models a non-trivial problem in itself. On the other hand, the likelihood must be integrated over the distribution of the random effects. Thus, numerical or analytical approximations are inevitable. The likelihood for SDE models can be approximated in various ways. Given discrete time observations, the likelihood is expressed in terms of the transition density. Approximation methods for the transition density reach from

solving the Fokker–Planck equation numerically (Lo, 1988), over standard first-order (Euler–Maruyama) or higher order approximation schemes and simulation-based approaches (Pedersen, 1995; Durham and Gallant, 2002) to a closed form approximation via Hermite polynomial expansion (Aït-Sahalia, 2002). If continuous time observations are assumed (e.g. if high frequency data are available), transition densities are not needed and the likelihood can be obtained from the Girsanov formula (Phillips and Yu, 2009). Popular analytical approximation techniques for general non-linear mixed effects models are first-order conditional estimation (Beal and Sheiner, 1981) and the Laplace approximation (Wolfinger, 1993). A computational alternative is the expectation–maximization algorithm, or stochastic versions thereof (Delyon *et al.*, 1999).

In the context of SDMEMs, the above-mentioned approximation methods have been combined in various ways, depending on whether observations are modelled in discrete or in continuous time (here we do not consider measurement noise). For discrete time observations, Hermite expansion of the transition density has been combined with Gaussian quadrature algorithms and Laplace's approximation (Picchini *et al.*, 2010; Picchini and Ditlevsen, 2011). Mixed effects that enter the diffusion coefficient were investigated in Delattre *et al.* (2015, 2018). The case of continuous time observations of a univariate SDMEM with Gaussian and mixture of Gaussian mixed effects entering the drift linearly was considered in Delattre *et al.* (2013, 2016) and Maitra and Bhattacharya (2018a).

Two aspects that are important in modelling biomedical data were not covered by these works: on the one hand, the inclusion of covariate information on both fixed and random effects. The only case which has previously been treated is the inclusion of covariates on random effects, but with no fixed effects (Maitra and Bhattacharya, 2018b); on the other hand, the theoretical investigations of estimators when the state process is modelled by a multivariate, time inhomogeneous and non-linear SDE. The lack of both in a model implies considerable restrictions for practitioners and the purpose of this paper is to fill this gap.

If the drift function is linear in the parameters, the standard asymptotic properties of the maximum likelihood estimator (MLE) in multi-dimensional, time homogeneous, non-linear SDMEMs can be shown by a natural extension of the proofs in Delattre *et al.* (2013). In particular, the model likelihood turns into a neat expression, and all remaining model complexities (multi-dimensionality of the state, non-linearity and covariates) are conveniently hidden in the sufficient statistics. The results in Delattre *et al.* (2013) on the discretization error which arises when continuous time statistics are replaced by their discrete time versions hold as well in the more complex model set-up. Their approach has, however, two drawbacks. The first is model related: it is assumed that observations are identically distributed, which impairs the inclusion of subject-specific covariate information. The other drawback is proof related: the regularity assumptions imposed are rather restrictive; for instance, the density of the random effects may not be smooth. More importantly, the proofs for multi-dimensional processes and parameters become tedious with long matrix calculations. This can be avoided by the more general approach which builds on L_2-differentiability and the local asymptotic normality property of a sequence of statistical models (Le Cam, 2012; Ibragimov and Has'minskii, 2013). Therefore, we approach the theoretical investigations from the more general local asymptotic normality perspective.

In regression models, the convergence of the average Fisher information (FI) is a standard assumption which facilitates the verification of MLE asymptotics considerably. We address this condition in the SDMEM set-up and point out the difficulties that arise here, when observations are not identically distributed.

The paper is structured as follows. Section 2 introduces the model framework and hypothesis testing. Moreover, we exemplify the framework with covariates for affine mixed effects. Section 3 is devoted to a simulation study in an example of the linear model, which is a submodel of

the model that is used for the electroencephalography data, and which is common in pharmacokinetics and is motivated by a recent study on selenium metabolism in humans (Große Ruse *et al.*, 2015). Here, we study finite sample and discretization bias of the estimation procedure and properties regarding hypothesis testing, where we investigate the effect of a drug treatment (as encoded by a covariate with levels *treatment* and *placebo*). We then apply the SDMEM framework to electroencephalography recordings of epileptic patients in Section 4, in a more general linear model with the purpose of investigating how channel interactions differ between non-seizure and seizure states. We then discuss our results and framework. Finally, we conclude with an investigation of asymptotic properties of the MLE, in particular for the linear model of main interest, and present some further technical details.

2. Maximum likelihood estimation for stochastic differential mixed effects models with covariates

This section considers parameter inference when observations are independent, but not necessarily identically distributed: a setting that naturally occurs when covariate information is included in the model formulation.

2.1. Model formulation

We consider N r-dimensional stochastic processes $X^i = (X_t^i)_{0 \leqslant t \leqslant T^i}$ whose dynamics are governed by the SDEs

$$\mathrm{d}X_t^i = F(X_t^i, D_t^i, \mu, \phi^i)\mathrm{d}t + \Sigma(t, X_t^i)\mathrm{d}W_t^i, \qquad 0 \leqslant t \leqslant T^i, \quad X_0^i = x_0^i, \quad i = 1, \ldots, N. \tag{1}$$

The r-dimensional Wiener processes $W^i = (W_t^i)_{t \geqslant 0}$ and the d-dimensional random vectors ϕ^i are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geqslant 0}, \mathbb{P})$, which is sufficiently rich to ensure independence of all random objects W^i and ϕ^i, $i = 1, \ldots, N$. The d-dimensional vectors ϕ^i, $i = 1, \ldots, N$, are the so-called *random effects*. They are assumed to be $\mathcal{F}_0$ measurable and have a common (usually centred) distribution which is specified by a (parameterized) Lebesgue density $g(\varphi; \vartheta)\mathrm{d}\varphi$. The parameter $\vartheta \in \mathbb{R}^{q-p}$ is unknown, as well as the *fixed effect* $\mu \in \mathbb{R}^p$. The combined parameter $\theta = (\mu, \vartheta)$ is the object of statistical inference and is assumed to lie in the parameter space Θ, which is a bounded subset of $\mathbb{R}^q$. The $D^i : [0, T^i] \to \mathbb{R}^s$ encode subject-specific covariate information and are assumed to be known. They can also encode a general time dependence, which not necessarily is subject specific. The functions $F : \mathbb{R}^{r+s+p+d} \to \mathbb{R}^r, \Sigma : [0, T] \times \mathbb{R}^r \to \mathbb{R}^{r \times r}$, with $T = \max_{1 \leqslant i \leqslant N} T^i$, are deterministic and known and the initial conditions x_0^i are r-dimensional random vectors. We assume standard regularity assumptions on the drift (including the D^i) and diffusion functions to assure

(a) existence and uniqueness of the solution to equations (1) and
(b) existence and good behaviour of the Radon–Nikodym derivative

$$\begin{aligned} q^i(\mu, \varphi) := q^i(\mu, \varphi; X^i) &= \frac{\mathrm{d}\mathbb{Q}^i_{\mu,\varphi}}{\mathrm{d}\mathbb{Q}^i_{\mu_0,\varphi_0}}(X^i) \\ &= \exp\Bigg\{ \int_0^{T^i} (F(X_s^i, D_s^i, \mu, \varphi) - F(X_s^i, D_s^i, \mu_0, \varphi_0))' \Gamma^{-1}(s, X_s^i)\mathrm{d}X_s^i \\ &\quad - \frac{1}{2} \int_0^{T^i} (F(X_s^i, D_s^i, \mu, \varphi) - F(X_s^i, D_s^i, \mu_0, \varphi_0))' \Gamma^{-1}(s, X_s^i)(F(X_s^i, D_s^i, \mu, \varphi) \\ &\quad + F(X_s^i, D_s^i, \mu_0, \varphi_0))\mathrm{d}s \Bigg\}, \end{aligned}$$

where $\Gamma = \Sigma\Sigma'$ and $\mathbb{Q}^i_{\mu,\varphi}$ is the distribution of X^i conditioned on an observed $\phi^i = \varphi$ (and μ_0 and φ_0 are fixed). The function q^i is the *conditional likelihood* for subject i given that we have observed the random effect $\phi^i = \varphi$. Therefore, the *unconditional likelihood* for subject i is $p^i(\theta) := p^i(\theta; X^i) = \int_{\mathbb{R}^d} q^i(\mu, \varphi)\, g(\varphi; \vartheta) \mathrm{d}\varphi$.

We observe X^i at time points $0 \leqslant t^i_0 < t^i_1 < \ldots < t^i_{n_i} = T^i$ and the inference task consists in recovering the 'true' underlying θ based on observations $X^i_{t^i_0}, \ldots, X^i_{t^i_{n_i}}$, $i = 1, \ldots, N$. We approach this inference task by first supposing that we have the entire paths $(X^i_t)_{0 \leqslant t \leqslant T^i}$, $i = 1, \ldots, N$, at our disposal. On the basis of these we derive the continuous time MLE and discretize it in a second step. The bias that is introduced by the discretization is investigated theoretically and by simulations.

2.2. Affine Gaussian mixed effects

In many applications the fixed and random effects enter the drift in an affine manner:

$$F(X^i_t, D^i_t, \mu, \phi^i) = A(X^i_t, D^i_t) + B(X^i_t, D^i_t)\mu + C(X^i_t, D^i_t)\phi^i. \tag{2}$$

An example of equation (2) is a widely used class of compartment models, which we illustrate in a simulation study in Section 3, and in our main application in Section 4, where we analyse electroencephalography data from epileptic patients. Likelihood-based inference then becomes explicit if the random effects are Gaussian distributed: $g(\varphi; \Omega) = \mathcal{N}(0, \Omega)(\varphi)$. The separation of μ and ϕ^i in equation (2) enables the modeller to impose random effects on only a selection of fixed effects. The conditional likelihood turns into the compact expression

$$q^i(\mu, \varphi) = \exp(\mu' U_{1i} - \tfrac{1}{2}\mu' V_{1i}\mu + \varphi' U_{2i} - \tfrac{1}{2}\varphi' V_{2i}\varphi - \varphi' Z_i \mu)$$

with the sufficient statistics

$$\left.\begin{aligned} U_{1i} &= \int_0^{T^i} B(X^i_s, D^i_s)' \Gamma^{-1}(s, X^i_s)\{\mathrm{d}X^i_s - A(X^i_s, D^i_s)\mathrm{d}s\}, \\ V_{1i} &= \int_0^{T^i} B(X^i_s, D^i_s)' \Gamma^{-1}(s, X^i_s) B(X^i_s, D^i_s)\mathrm{d}s, \\ U_{2i} &= \int_0^{T^i} C(X^i_s, D^i_s)' \Gamma^{-1}(s, X^i_s)\{\mathrm{d}X^i_s - A(X^i_s, D^i_s)\mathrm{d}s\}, \\ V_{2i} &= \int_0^{T^i} C(X^i_s, D^i_s)' \Gamma^{-1}(s, X^i_s) C(X^i_s, D^i_s)\mathrm{d}s, \\ Z_i &= \int_0^{T^i} C(X^i_s, D^i_s)' \Gamma^{-1}(s, X^i_s) B(X^i_s, D^i_s)\mathrm{d}s. \end{aligned}\right\} \tag{3}$$

Integration over φ gives the unconditional likelihood for subject i:

$$p^i(\theta) = \frac{1}{\sqrt{\det(I + V_{2i}\Omega)}} \exp\left[\{U'_{1i} - U'_{2i} R^i(\Omega) Z_i\}\mu - \frac{1}{2}\mu'\{V_{1i} - Z'_i R^i(\Omega) Z_i\}\mu + \frac{1}{2} U'_{2i} R^i(\Omega) U_{2i}\right], \tag{4}$$

with $R^i(\Omega) = (V_{2i} + \Omega^{-1})^{-1}$. In particular, the MLE $\hat{\mu}_N$ of the fixed effect (given Ω) is explicit:

$$\hat{\mu}_N(\Omega) = \left[\sum_{i=1}^N \{V_{1i} - Z'_i R^i(\Omega) Z_i\}\right] \sum_{i=1}^N \{U_{1i} - Z'_i R^i(\Omega) U_{2i}\}. \tag{5}$$

Remark 1. The likelihood p^i is explicit even if the fixed effect enters the drift non-linearly. However, only a linear fixed effect μ leads to an explicit expression for its MLE.

2.2.1. Discrete data

Above we assumed that we observe the entire paths $(X_t^i)_{0\leqslant t\leqslant T}$. In practice, observations are available only at discrete time points $t_0,\dots,t_n$. A natural approach is to replace the continuous time integrals in $q^i(\theta)$ by discrete time approximations and to derive an approximate MLE based on the resulting approximate likelihood. For instance, an expression of the form $\int_{t_k}^{t_{k+1}} h(s,X_s^i)\mathrm{d}X_s^i$ may be replaced by a first-order approximation $h(t_k,X_k^i)\Delta X_k^i$. In the linear model (2), the approximation of the continuous time likelihood corresponds to the exact likelihood of its Euler scheme approximation. In particular, if we observe all individuals at time points $t_k = Tk/n$ and denote by $U_{1i}^n, V_{1i}^n, U_{2i}^n, V_{2i}^n$ and Z_i^n the first-order discrete time approximations to the continuous time statistics $U_{1i}, V_{1i}, U_{2i}, V_{2i}$ and Z_i in equations (3), we have the following result.

Theorem 1 (negligibility of discretization error). Assume model (2) and suppose that A, $B'\Gamma^{-1}B$, $B'\Gamma^{-1}C$, $C'\Gamma^{-1}C$, $B'\Gamma^{-1}$ and $C'\Gamma^{-1}$ are globally Lipschitz continuous in (t,x) and that in addition to A, B, C and Σ also $B'\Gamma^{-1}$ and $C'\Gamma^{-1}$ are of sublinear growth in x, uniformly in t. Then, for all $p\geqslant 1$ and all $i=1,\dots,N$, there is a constant C such that

$$\mathbb{E}_{\theta_0}(\llbracket V_{1i}-V_{1i}^n\rrbracket^p + \|U_{1i}-U_{1i}^n\|^p + \llbracket V_{2i}-V_{2i}^n\rrbracket^p + \|U_{2i}-U_{2i}^n\|^p + \|Z_i-Z_i^n\|^p) \leqslant C\left(\frac{T}{n}\right)^{p/2}.$$

The discretization error is investigated numerically in Section 3.

2.3. Hypothesis testing

It is commonly of interest to test whether an applied treatment has a significant effect on the treated subjects, i.e. to test whether an underlying treatment effect β, an l-dimensional subparameter of the fixed effect μ, $1\leqslant l\leqslant p$, is significantly different from 0. The asymptotic normality of the MLE for the SDMEMs lends itself naturally to the application of Wald tests, which can be used to investigate two-sided null hypotheses such as $H_0:\beta=0$ (no treatment effect) or, more generally, any k-dimensional, $1\leqslant k\leqslant l$, linear null hypothesis $H_0: L\beta=\eta_0$, where L is a $k\times l$ matrix of rank k, specifying the linear hypotheses of interest, and $\eta_0\in\mathbb{R}^k$. The Wald test statistic is $(L\hat{\beta}_N-\eta_0)'(L\hat{V}_N L')^{-1}(L\hat{\beta}_N-\eta_0)$, where $\hat{\beta}_N$ is the MLE of β and $\hat{V}_N=\widehat{\mathrm{cov}}(\hat{\beta}_N)$ denotes the estimated variance–covariance matrix of $\hat{\beta}_N$. Under the null hypothesis the test statistic is asymptotically χ^2 distributed with k degrees of freedom (Lehmann and Romano, 2006). Alternatively, the likelihood ratio test can be applied. Let p_0 and p_a denote the likelihoods under the null and under the alternative; then the test statistic $-2\log(p_0/p_a)$ is asymptotically χ^2 distributed with degrees of freedom equal to the difference in number of parameters. Hypothesis testing in the present SDMEM framework will be further illustrated in the following two sections.

3. Simulation study on the linear transfer model

The model under investigation provides a proof of concept of our main application of modelling electroencephalography measurements of epileptic patients, which will be done in the next section, but is also inspired from a study on selenomethionine metabolism in humans (Große Ruse *et al.*, 2015). This multi-dimensional linear transfer model finds frequent applicability in pharmacokinetics. Each component of the model's state vector represents the concentration of a substance in a certain compartment (e.g. in an organ of the human body), such that the model describes the flow between several compartments. We consider a flow in the form of a cascade-

shaped transfer structure, illustrated in Fig. 1. The transfer rates between compartments are mostly subject specific, which we account for by inclusion of random effects. In this way, the total variation is split into within- and between-individual components. Additionally, dynamics may depend on covariates D^i. Here, the $D^i \in \{0, 1\}$ encode the randomly assigned treatment group of subject i. For simplicity we assume a unit diffusion matrix, such that we consider the model

$$\mathrm{d}X_t^i = F(X_t^i, D^i, \mu, \phi^i)\mathrm{d}t + \mathrm{d}W_t^i = -G(\alpha + \phi^i)X_t^i + D^i\beta\mathrm{d}t + \mathrm{d}W_t^i,$$

for $0 \leqslant t \leqslant T$ and $X_0^i = 0$, where $\mu' = (\alpha', \beta')$ is the fixed parameter and

$$G(\alpha) = \begin{pmatrix} \alpha_1 & 0 & 0 & 0 & -\alpha_5 \\ -\alpha_1 & \alpha_2 & 0 & 0 & 0 \\ 0 & -\alpha_2 & \alpha_3 + \alpha_6 & 0 & 0 \\ 0 & 0 & -\alpha_3 & \alpha_4 & 0 \\ 0 & 0 & 0 & -\alpha_4 & \alpha_5 \end{pmatrix}.$$

This is a special case of the affine model (2). For a given random effect ϕ^i, this is an Ornstein–Uhlenbeck (OU) process, restricted to have unit diffusion and certain entries in the drift matrix equal to 0. We shall discuss more details in the next section, when treating the electroencephalography data. The (unknown) fixed effect μ has the six-dimensional component α, which is shared across both groups (*placebo* and *treatment*) and an additional five-dimensional component β, which describes the effect of the covariate (treatment) on a subject's dynamics. The fixed effects provide information about the entire population. We let $\beta' = (1, 2, 3, 1, -2)$. The random effects ϕ^i are independent and identically distributed $\mathcal{N}(0, \Omega)$ with unknown Ω, which we for simplicity and to avoid overparameterization assume is diagonal with entries $\mathrm{diag}(\Omega) = (0.5^2, 1^2, 1^2, 0.5^2, 0.3^2, 0.3^2)$. The random effects quantify how the dynamics of a specific subject differ from the population. The variance parameters in Ω govern the between-subject variability, whereas the Wiener processes govern the within-subject variability. If the between variability is large compared with the within variability, i.e. subjects follow quantitatively different dynamics, it is important to include the random effects for robust statistical estimation; for example, we gain power to detect a possible treatment effect. With $\alpha' = (\alpha_1, \ldots, \alpha_6) = (2, 4, 3, 2, 1, 1)$, all eigenvalues of $G(\alpha)$ have positive real parts, implying that the model has a stationary solution. The processes X^i for individuals without treatment, $D^i = 0$, are (on average) mean reverting to 0, whereas those for individuals in the treatment group have average long-term mean

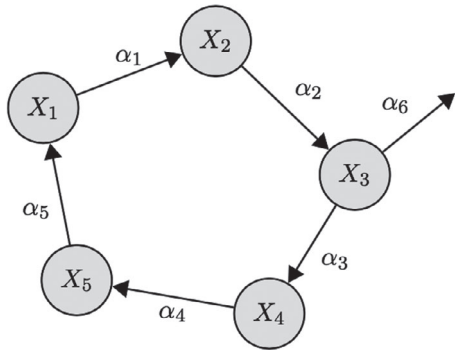

Fig. 1. Illustration of the five-dimensional linear transfer model used in the simulation example: the state $X_j = (X_{j,t})_{0 \leqslant t \leqslant T}$ gives the concentration (over time) of a substance in compartment j, $j = 1, \ldots, 5$; the α_j, $j = 1, \ldots, 5$, are the unknown flow rates between compartments and α_6 represents the outflow rate of the system

$G(\alpha)^{-1}\beta = (7.50, 4.25, 5.00, 8.00, 14.00)'$; see also Fig. 2. The observation horizon is fixed to $T = 15$. A trajectory of $(X_t^1, \ldots, X_t^N)_{0 \leqslant t \leqslant T}$ is simulated with the Euler–Maruyama scheme with simulation step size $\delta = 10^{-4}$. Fig. 2 shows four realized trajectories of the five-dimensional process X^i. Figs 2(a) and 2(b) show trajectories for $D^i = 0$ and the lower two correspond to $D^i = 1$.

3.1. Parameter estimation

To mitigate simulation errors, the simulated trajectories are thinned by a factor b (taking only every bth observation). We explore the expected time discretization bias of the estimators by repeating estimation for different thinning factors $b \in \{10, 100\}$, which results in sampling intervals $\Delta t = \delta b = 0.001, 0.01$. To investigate estimation performance as a function of sample size, we used $N = 20$ and $N = 50$. Estimation for the $(\Delta t, N)$ combinations considered was repeated on $M = 500$ simulated data sets. Table 1 reports the sample estimates of relative biases and root-mean-squared errors (RMSEs) of the fixed effects and of the variances of the random effects. The relative bias of $\hat{\alpha}_j$ is computed as $(1/M)\Sigma_{m=1}^{M}(\hat{\alpha}_j^{(m)} - \alpha_j)/\alpha_j$ and the RMSE as $\{(1/M)\Sigma_{m=1}^{M}(\hat{\alpha}_j^{(m)} - \alpha_j)^2\}^{1/2}$, $j = 1, \ldots, 6$, with an analogous definition for the other parameters. The first six rows in Table 1 correspond to estimated biases and RMSEs of the shared fixed effects α_j, $j = 1, \ldots, 6$. The subsequent five rows show these metrics for the treatment effects β_j, $j = 1, \ldots, 5$, and the last six rows correspond to the metrics for the diagonal elements of Ω (i.e. the variances of the random effects). The estimation is very accurate already at sample sizes as small as $N = 20$, when the data are sampled at high frequency (here $1/0.001$). Increasing the sample size to $N = 50$ does not add much to the accuracy of the estimation of the fixed effects. But it does, and not surprisingly, improve the estimation of the variances of the random effects, by up to 14 percentage points. For a lower sampling frequency of $1/0.01$, estimates of the fixed effects α and β are on average biased by only about 1–2%, which is still very accurate. The variances of the random effects are estimated with an average bias of 5–9% for $N = 50$ and $\Delta t = 0.01$. Not displayed here are simulation results for low frequency observations with $\Delta t = 0.1$. As predicted by theorem 1, simulations show that estimation becomes fairly unreliable in this case. The bias due to the time discretization of the continuous time estimator is pronounced, with values of up to 25% for the fixed effects and up to almost 50% for the variances of the random effects. The RMSEs rise by more than 100%, compared with the results that were obtained for a 10-times higher sampling frequency. If only low frequency data are available, caution is therefore recommended and estimation should only be done on a data set that has been enlarged by imputing data in between the observation time points.

3.2. Hypothesis testing

A natural step is to test whether the treatment effect β, or a subparameter, is significantly different from 0. We estimate the false positive rate of the Wald test (see Section 2.3) for this model and investigate the test's power under various non-zero treatment effects. The estimated variance–covariance matrix $\hat{V}_N = \widehat{\mathrm{cov}}(\hat{\beta}_N)$ of $\hat{\beta}_N$ is obtained from $M = 500$ (separately) computed MLEs $\hat{\beta}_N^{(m)}$, $m = 1, \ldots, M$, where underlying data sets have been simulated under the true hypothesis (under H_0 for estimation of the false positive rate and under H_1 for power estimation). Table 1 shows that the estimation was accurate for high and medium frequency observations. Diagnostic plots (which are not shown here) reveal that the asymptotic distribution of the MLE is close to normal already for $N = 20$ subjects, such that, even for a rather small data set and a medium sampling frequency, test results are reliable. The choice $(N, \Delta t) = (20, 0.01)$ provides a simulation setting which is sufficiently reliable, but at the same time not trivial and will challenge the hypothesis test, in particular for small treatment effects. The estimated false positive rate

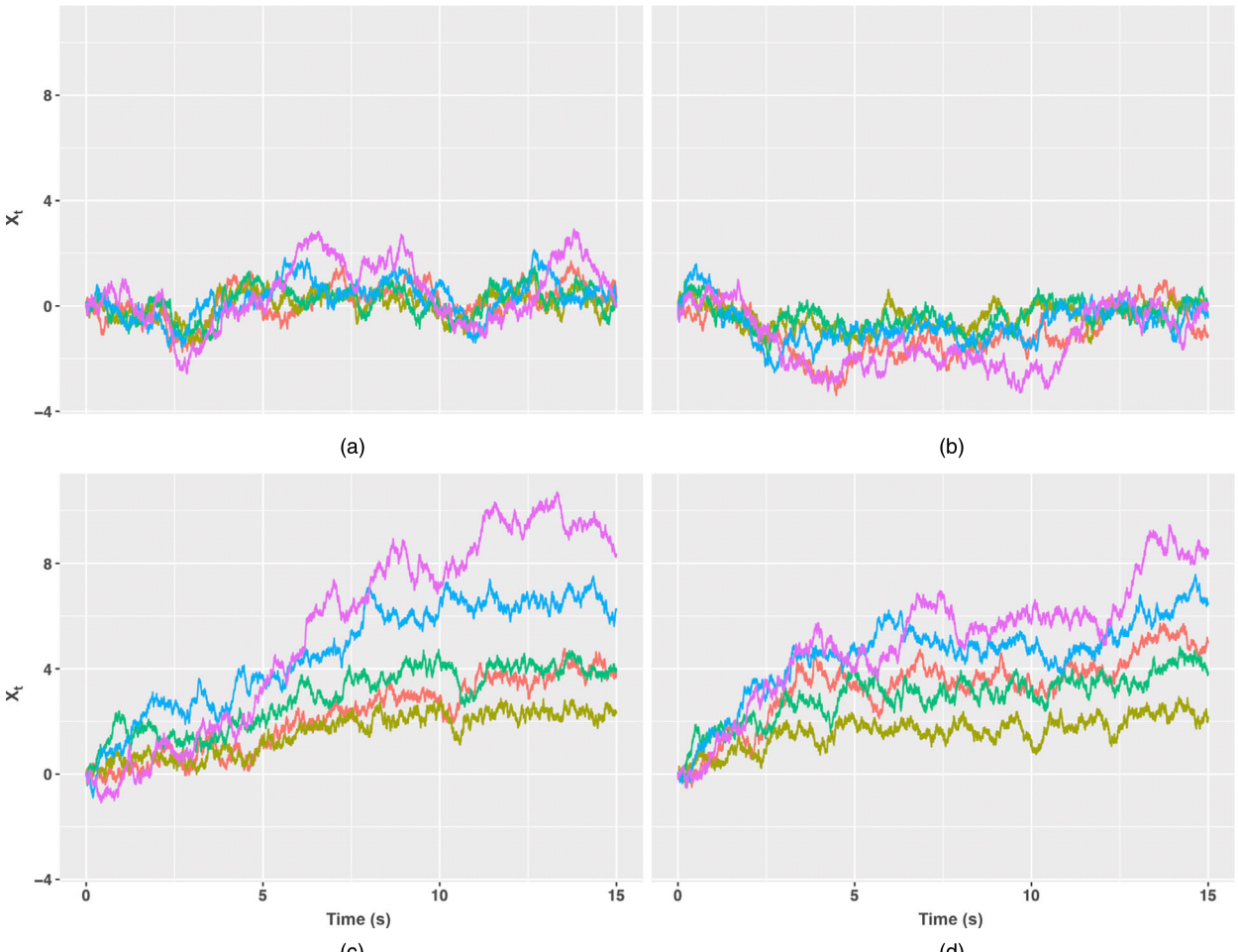

Fig. 2. Linear transfer model: four realizations of the five-dimensional state process (——, X^1; ——, X^2; ——, X^3; ——, X^4; ——, X^5): (a), (b) realizations when the covariate is 0 (reference group); (c), (d) trajectories for $D^i = 1$ (treatment group); note the clearly visible difference in the long-term means between the two groups

Table 1. Linear transfer model†

True value	*Results for* $(N, \Delta t) = (20, 0.001)$		*Results for* $(N, \Delta t) = (50, 0.001)$		*Results for* $(N, \Delta t) = (50, 0.01)$	
	Relative bias	*RMSE*	*Relative bias*	*RMSE*	*Relative bias*	*RMSE*
α						
2.00	0.003	0.116	0.001	0.079	−0.018	0.086
4.00	0.001	0.232	−0.002	0.149	−0.024	0.172
3.00	0.003	0.253	0.001	0.163	−0.021	0.170
2.00	−0.003	0.126	−0.001	0.083	−0.017	0.088
1.00	0.003	0.074	0.001	0.047	−0.016	0.049
1.00	−0.003	0.146	0.002	0.091	−0.008	0.091
β						
1.00	0.000	0.157	−0.002	0.099	−0.020	0.099
2.00	−0.001	0.174	−0.002	0.114	−0.024	0.121
3.00	0.002	0.233	0.002	0.152	−0.010	0.152
1.00	0.010	0.231	−0.001	0.148	0.014	0.146
−2.00	0.006	0.203	0.002	0.124	−0.024	0.131
diag(Ω)						
0.25	−0.091	0.093	−0.037	0.062	−0.079	0.062
1.00	−0.046	0.355	−0.035	0.208	−0.095	0.216
1.00	−0.073	0.343	−0.035	0.215	−0.085	0.219
0.25	−0.035	0.097	−0.026	0.061	−0.065	0.060
0.09	−0.045	0.035	−0.009	0.022	−0.047	0.021
0.09	−0.181	0.055	−0.040	0.036	−0.065	0.035

†Shown are the estimated relative bias and RMSE of $\hat{\alpha}, \hat{\beta}$ and diag$(\hat{\Omega})$. The sample sizes are $N = 20, 50$, and the sampling intervals are $\Delta t = 0.001, 0.01$. For every combination $(N, \Delta t)$, the estimation was repeated on $M = 500$ generated data sets.

(based on M under H_0 generated data sets) is 0.074, revealing a slightly liberal finite sample test behaviour. The power of detecting a treatment effect (rejecting $H_0 : \beta = 0$) was computed for various values of β. For $\beta = (1, 2, 3, 1, -2)'$ (values as above), the estimated power was 1. This comes as no surprise as the long-term mean $(7.5, 4.25, 5, 8, 14)'$ of the state process in the treatment group is considerably different from the zero long-term mean of the control group. The power, estimated as 0.956, was still convincing for a much smaller treatment effect $\beta = (0.1, 0.2, 0.3, 0.1, -0.2)'$, which gives a long-term mean of $(0.75, 0.425, 0.5, 0.8, 1.4)'$. This is especially impressive as the state process's standard deviation (from its long-term mean 0) under H_0 is about $(0.66, 0.49, 0.59, 0.72, 1.21)'$. More challenging is the rejection of H_0 when the treatment has a small effect on, for example, only one co-ordinate, $\beta = (0.1, 0, 0, 0, 0)'$. In this case (long-term mean $(0.2, 0.1, 0.1, 0.15, 0.3)'$), and for $N = 20$, the chance of rejecting H_0 is as small as 16% and it is thus hardly possible to detect a difference between groups. However, although being only slightly conservative, the asymptotic Wald test can detect a treatment effect for a rather small data set, even if it causes only a little change of the long-term mean compared with the standard deviation of the process.

4. Analysis of electroencephalography data

Scalp electroencephalography is a non-invasive method to measure electrical activity in the brain over time, recorded by electrodes placed on the scalp. Abnormal patterns in the recorded brain

waves are used as possible indicators for diseases such as epilepsy and can help in determining a suitable treatment for the patient. The data set was collected during a study that was conducted by the Children's Hospital Boston and is described in Shoeb (2009). It consists of continuous electroencephalography recordings on 23 epilepsy patients. The electrodes were arranged on the scalp according to the international 10–20 system and the electroencephalography signal was recorded with a sampling frequency of 256 Hz. This is high frequency compared with the typical timescales of the system, and thus, for this type of data, the discretization error will be negligible. During the time of recording, every patient experienced one or more periods of abnormal activity that have been classified as epileptic seizures by Shoeb (2009).

Part of this data set was also analysed in Østergaard *et al.* (2017). Their results, which were obtained by using a different modelling approach, indicated increased channel interaction strength during seizure. However, their findings were based on data from a single subject only. It is therefore of interest whether we can infer an increased interaction when combining data from several subjects within a dynamical mixed effects framework. We focus our analysis on recordings from four channels in the frontal lobe, FP1_F7, FP1_F3, FP2_F4 and FP2_F8, as done in Østergaard *et al.* (2017). Thus, responses are four-dimensional time series for every patient. The first two channels are on the left hemisphere and the second two are, mirrored, on the right. For every patient we extracted two 5-s periods of recording, one of them reflecting normal brain activity and the other reflecting abnormal activity classified as epileptic seizure. Fig. 3 shows data for the selected periods of two subjects.

The dynamics of the signals during seizure differ clearly from preseizure behaviour and the objective of this analysis is to understand better, quantitatively and qualitatively, how they differ. From a neurophysiological viewpoint the interaction structure between brain regions or different channels is of interest and, in particular, if and how this network structure changes under different conditions, e.g. when patients enter an epileptic seizure state. A hint on possible interactions can be obtained by investigating the correlation structure between channels. Under a sufficiently short time window, the otherwise non-stationary behaviour of spontaneous brain activity can be considered stationary. We model the 5-s sections of electroencephalography recordings from the four selected channels with an OU process, as in the simulation study, now in four dimensions and with no imposed 0s. This is a process with dynamics $\mathrm{d}X_t = AX_t\mathrm{d}t + \Sigma\mathrm{d}W_t$ and explicit solution $X_t = \exp(At)x_0 + \int_0^t \exp\{A(t-s)\}\Sigma\mathrm{d}W_s$. In particular, X_t (given x_0) is Gaussian with mean $\mathbb{E}(X_t) = \exp(At)x_0$ and covariance matrix $\mathbb{V}(X_t) = \int_0^t \exp(As)\Sigma\Sigma'\exp(A's)\mathrm{d}s$. Thus, A corresponds to $-G$ of the simulation study. If all eigenvalues of the rate matrix A have negative real parts, X has a stationary solution and the stationary distribution is a centred Gaussian distribution with covariance matrix $V = \int_0^\infty \exp(Au)\Sigma\Sigma'\exp(A'u)\mathrm{d}u$ and auto-correlation function $r_X(\tau) = V^{1/2}\exp(A'\tau)V^{-1/2}$.

4.1. The statistical model

The prevalent intersubject variability for electroencephalography data is one of the greater challenges for any inference procedure (Shoeb, 2009), and we account for such subject-specific deviations from mean OU dynamics by the inclusion of random effects. We present the subject-specific SDMEM model for the electroencephalography data first and afterwards give a motivation for our choice. We denote the preseizure process of subject i by $Y^{i,1}$ and the seizure process by $Y^{i,2}$. During seizure, the signal is amplified considerably (Fig. 3). As structural differences are easier to analyse when preseizure and seizure data are of comparable magnitude, we rescale the data to $X_t^{i,k} = \mathrm{diag}(1/\sigma_{11}^{i,k}, \ldots, 1/\sigma_{44}^{i,k})Y_t^{i,k}$, with $\sigma_{jj}^{i,k}$ being the infinitesimal standard deviation (the square root of the quadratic variation) of channel j. Normalizing by a diagonal matrix

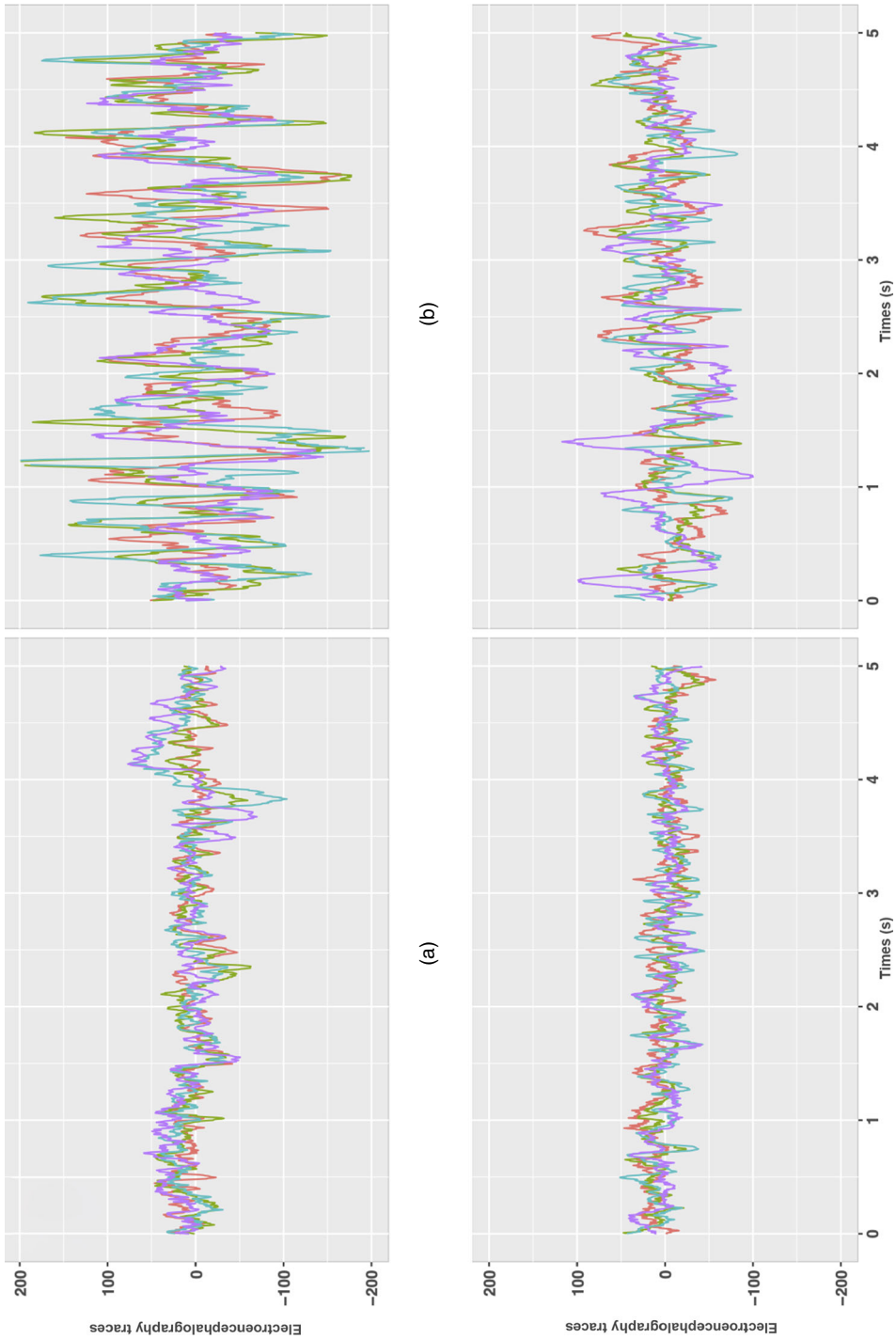

Fig. 3. Sections of 5 s of the electroencephalography recording during preseizure activity and during an epileptic seizure for the four channels FP1_F7 (——), FP1_F3 (——), FP2_F4 (——) and FP2_F8 (——) for two subjects: (a) subject 1 (male; 11 years old), preseizure; (b) subject 1, seizure; (c) subject 2 (male, 22 years old), preseizure; (d) subject 2, seizure

does not introduce changes in the inherent channel structure but only affects the scaling. The specific choice of the scaling renders the quadratic variation of the obtained processes $X^{i,k}$ to be a correlation matrix type. Taking the OU dynamics as the base model, we then model the (normalized) data for subject i by

$$\mathrm{d}X_t^{i,k} = \{A + \Phi^{i,1} + D^{i,k}(M + \Phi^{i,2})\}X_t^{i,k}\mathrm{d}t + \Sigma \mathrm{d}W_t^{i,k}, \tag{6}$$

where $W^{i,k}$ are independent Brownian motions, $A, M, \Phi^{i,1}$ and $\Phi^{i,2}$ are 4×4 matrices and the entries of $\Phi^{i,1}$ and $\Phi^{i,2}$ are independent centred Gaussian random variables (the random effects). Note that the random effect $\Phi^{i,1}$ is active during both preseizure and seizure, whereas $\Phi^{i,2}$ is activated only during seizure and can be interpreted as a nested random effect. The covariate $D^{i,k}$ encodes whether the data belong to the preseizure ($D^{i,1} = 0$) or seizure state ($D^{i,2} = 1$). Thus, for a preseizure state, population dynamics are driven by the rate matrix A, whereas M represents the covariate (or seizure) effect. Rewriting equation (6) as

$$\mathrm{d}X_t^{i,k} = \left\{ B(X_t^{i,k}, D^{i,k})\mu + C(X_t^{i,k}, D^{i,k}) \begin{pmatrix} \phi^{i,1} \\ \phi^{i,2} \end{pmatrix} \right\}\mathrm{d}t + \Sigma \mathrm{d}W_t^{i,k} \tag{7}$$

(with $\phi^{i,1}$ and $\phi^{i,2}$ being the vectorized versions of $\Phi^{i,1}$ and $\Phi^{i,2}$ respectively) reveals that this model belongs to the class of affine SDMEMs with covariates, model (2), and thus has explicit likelihood and fixed effects estimators.

An alternative to the random-effects model is to make separate analyses on each subject, and then to summarize the results, for example by taking averages over parameters from different subjects, to obtain estimates for the population. This approach has some drawbacks. Firstly, the data are not fully used and some statistical power is lost; secondly, it can be difficult to evaluate whether averages over individual parameter estimates are the correct measures for the population. These will depend on the parameterization, and maybe the average of some non-linear transformation is more appropriate. Below, we also include this analysis to compare with the random-effects approach.

4.2. Motivation for the model approach

The processes $W^{i,1}$ and $W^{i,2}$ represent the noise within the system on a short timescale. Their independence is supported by the fact that data sections $X^{i,1}$ and $X^{i,2}$ are temporally (on a larger timescale) clearly separated. In general, behaviour during seizures is more variable, and, in particular, shows a stronger amplification.

Fig. 4 shows that the average structure of the infinitesimal correlations between channels (off-diagonal plots) does not differ considerably between the preseizure (left-hand side) and seizure states (right-hand side). The estimated infinitesimal standard deviations $\hat{\sigma}_{jj}^{i,k}$ of the channels (diagonal plots) reveal, however, that in most subjects and channels (80%) the standard deviation increases; in the most extreme case it increases 14 fold, and in 78% of the cases it more than doubles. Because of the shared infinitesimal correlation structure we model the normalized preseizure and seizure processes with the same diffusion matrix, denoted above by Σ. This implies that any further changes apart from the scaling in the dynamics between states are captured by changes in the drift. The transition from the preseizure to seizure state is modelled in terms of the drift matrix $M + \Phi^{i,2}$. The structural change in the population dynamics is represented by M, and the change in the subject-specific variation due to seizure is represented by the random effect $\Phi^{i,2}$.

4.3. Results

The statistical conclusions are based on the population rate matrices A and M, which are

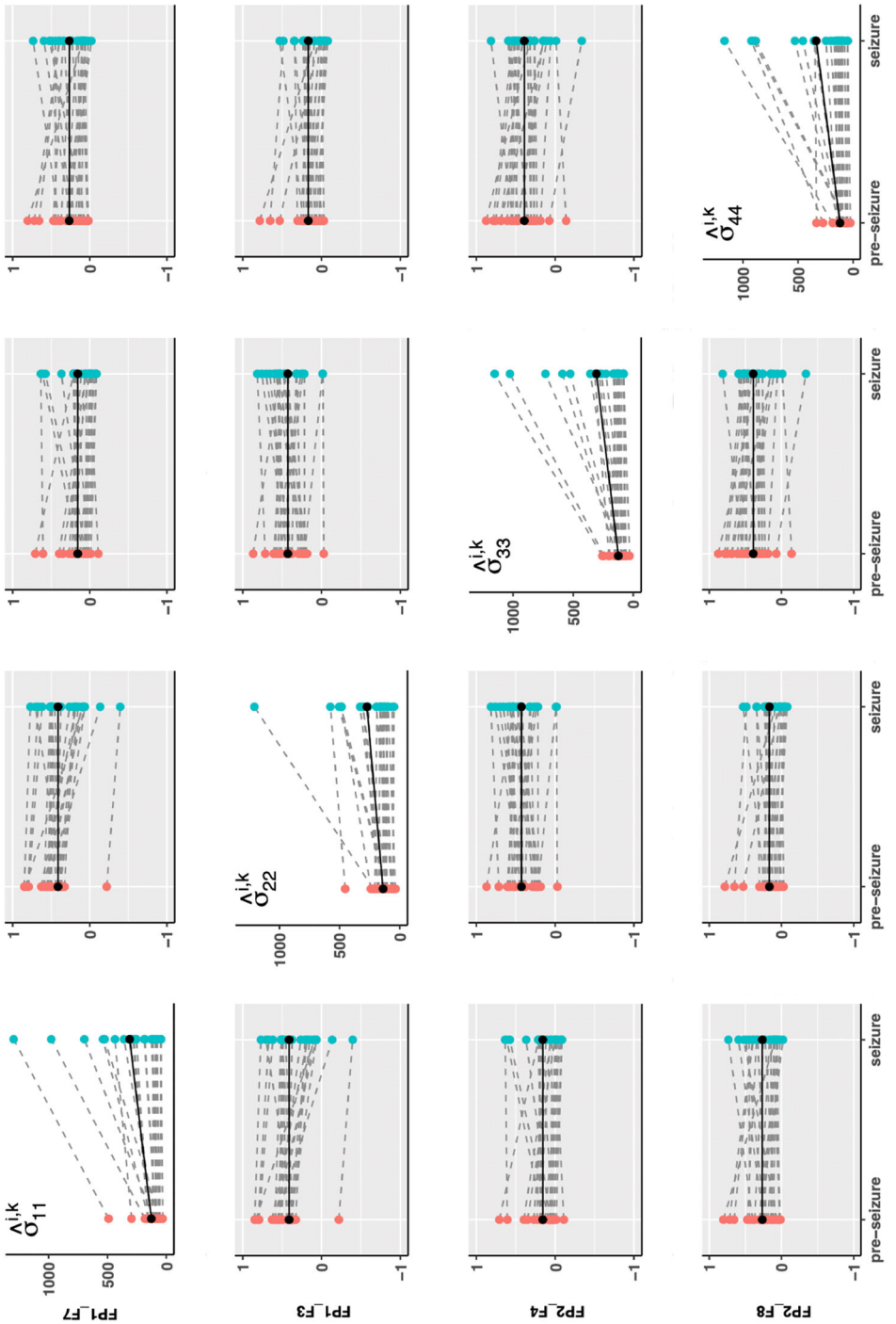

Fig. 4. Diagonal plots, infinitesimal standard deviation for every channel, estimated by the square root of the quadratic variation, used to normalize the data records; off-diagonal plots, infinitesimal correlation between channels (– – – –), individual estimates for preseizure (left-hand side) and seizure (right-hand side) data; ———, mean correlation, averaged over individuals and states): (a) FP1_F7; (b) FP1_F3; (c) FP2_F4; (d) FP2_F8

estimated by their MLEs as outlined in Section 2.2. The estimates of the population-based rate matrices are

$$\hat{A} = \begin{pmatrix} -10.52 & -3.59 & -0.42 & 2.47 \\ 3.24 & -17.72 & 4.76 & 1.70 \\ 1.98 & 0.14 & -12.60 & 3.94 \\ 0.74 & -1.75 & -1.52 & -12.87 \end{pmatrix},$$

$$\hat{M} = \begin{pmatrix} -3.22 & 2.65 & 0.80 & -0.16 \\ 0.83 & 4.60 & -1.51 & 2.81 \\ -0.82 & -0.27 & 0.74 & 0.00 \\ 3.27 & 0.59 & 1.30 & -3.36 \end{pmatrix}.$$

The eigenvalues of $\hat{A}$ and $\hat{A} + \hat{M}$ have negative real parts, such that stationary distributions on the population level for the preseizure and seizure states indeed exist.

In a first step we assess whether the overall covariate effect M is significant by testing $H_0 : M = 0$ *versus* $H_0 : M \neq 0$ with a likelihood ratio test. The likelihood ratio statistic, which under H_0 is asymptotically χ^2_{32-16} distributed, has a realized value of 13.71, with a p-value of 0.62. We conclude that the null hypothesis $H_0 : M = 0$ cannot be rejected on a 5% level. However, the data set consists of observations from only 23 subjects and the number of fixed effects alone (32 parameters) is considerably higher. Therefore, a possible prevalent covariate effect is difficult to detect. More insight into where changes might be present in the rate matrix between preseizure and seizure states is provided in Fig. 5. It shows the 95% confidence intervals (CIs) for every entry of M in blue. Only one element of M has a CI that does not include 0. A way to increase statistical power is to cut down on the number of unknown parameters. Considering only one element of M active instead of all 16, the number of unknown fixed effects is reduced from 32 to 17. Each of the black CIs in Fig. 5 is derived from a reduced model in which all except the one element of M being tested are set to 0. As expected, most CIs are more narrow; however, only a few elements appear to have an effect. The lower left-hand plot, for example, suggests an increased influence of channel FP2_F8 on FP1_F7 under seizure.

The analysis was repeated on each subject individually to evaluate the importance of including random effects. Taking averages over the 23 individual estimates, the following population estimates of the rate matrices were obtained:

$$\hat{A} = \begin{pmatrix} -10.69 & -4.99 & 0.25 & 1.46 \\ 4.06 & -20.32 & 5.98 & 0.51 \\ 2.39 & -0.65 & -13.17 & 4.14 \\ 0.25 & -1.47 & -1.83 & -13.55 \end{pmatrix};$$

$$\hat{M} = \begin{pmatrix} -3.41 & 4.04 & 0.38 & 0.75 \\ -0.01 & 6.85 & -3.01 & 4.53 \\ -1.26 & 0.23 & 0.94 & 0.08 \\ 4.49 & 0.44 & 1.68 & -3.46 \end{pmatrix}.$$

The rate matrices are similar to the rate matrices that were estimated from the model with random effects; however, the variances of the estimators are larger because of the less efficient use of the data. This is illustrated in Fig. 5 which includes the individual estimates in pink from analyses on each subject, which clearly shows the large between-subjects variability. Moreover, in the individual analyses outlier estimates appear, which for readability are not included in Fig. 5. This shows that performing individual analyses on each subject leads to more volatile estimates. These estimates are then summarized in the red CIs, which are the estimates of

Fig. 5. Confidence intervals for the estimated entries of the covariate effect matrix M (——, full model with random effects; ——, derived from 16 reduced models where all except one entry of M are set to 0; ——, individual estimates for each subject, obtained from individual analyses; ——, CIs of the averages from these individual analyses on each subject (the x-axes have been truncated for readability; between 0 and 3 outliers are outside the plotting areas in each subplot, the maximum absolute values being around 100): (a) FP1_F7; (b) FP1_F3; (c) FP2_F4; (d) FP2_F8

Table 2. Stationary correlations between channels, for the preseizure state and the seizure state†

Channel 1	*Channel 2*	*Correlation*		*Change (%)*
		Preseizure	*Seizure*	
FP1_F7	FP1_F3	0.42	0.52	23.81
FP1_F7	FP2_F4	0.22	0.26	18.18
FP1_F7	FP2_F8	0.32	0.43	34.37
FP1_F3	FP2_F4	0.60	0.59	−1.67
FP1_F3	FP2_F8	0.23	0.36	56.52
FP2_F4	FP2_F8	0.43	0.47	9.30

†The last column shows the change in correlation for seizure epochs compared with non-seizure periods.

the means of the subject-specific estimates. This is the usual approach to estimate population parameters, when random effects are not taken into account. We see how we lose statistical power with this approach, where CIs are broader, not using the full information of the data. Note that the individual analyses require more parameters, namely one for each subject for each rate parameter, whereas the random-effects model requires only one variance parameter for each rate parameter. Finally, the random-effects model is more readily interpretable in terms of generalizing to similar patients outside the study.

It is not straightforward to interpret a covariate effect by looking at the matrix M entry by entry. Insights about structural changes in the underlying dynamics can more easily be gained by looking at interactions in the system. Interactions can be assessed by the correlations between components of $X_t^{i,k}$. To analyse this, we compare the (population) stationary covariance matrices of the preseizure and seizure state, which will reveal differences in the long-run correlation structure between channels. The population estimates of the correlation matrices of the stationary distributions for preseizure and seizure states are shown in Table 2. In line with the findings in Østergaard *et al.* (2017), channel correlations increase during seizure, most of them by at least around 20%.

Other quantities of interest are the auto-correlation functions that are shown in Fig. 6. The diagonal panels in Fig. 6 show the univariate auto-correlation for every channel, i.e. the correlations between a channel and its time-lagged version, as a function of the time lag. The auto-correlations show no marked difference between the preseizure and seizure states. This can also be summarized by the eigenvalues of matrices $\hat{A}$ and $\hat{A} + \hat{M}$. The absolute values of the real parts provide the rates of decay and, thus, their inverses indicate the typical time constants in the system. For the preseizure state the absolute values of the real parts vary between 11.4 and 17.7, whereas during seizure these vary between 11.6 and 18.0.

To summarize, despite not being statistically significant, there are indications of changes in the correlation structures during epileptic seizures, where correlations between channels increase. Nonetheless, the main effect of an epileptic seizure seems to be captured by increased variance addressed in the rescaling of the data, more than structural changes. However, with only 23 patients finer effects might be difficult to unravel. An analysis of all channels would be of interest but is only possible with much larger sample sizes because of the many parameters that a full analysis would imply.

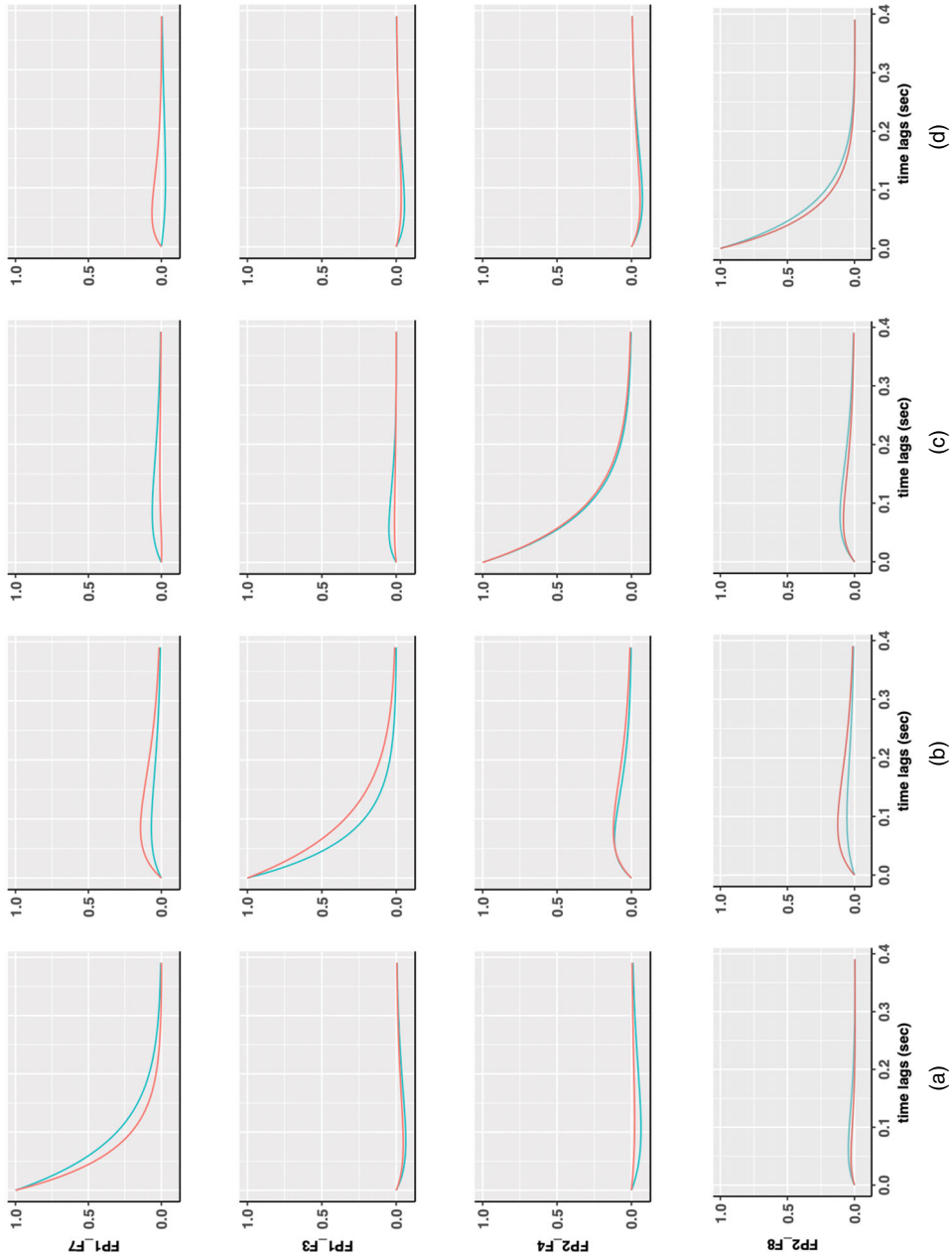

Fig. 6. Theoretical multivariate auto-correlation functions $r^k_{\hat{X}}(\tau) = \hat{V}_k^{1/2} \exp(\hat{A}'_k \tau) \hat{V}_k^{-1/2}$, $k = 1, 2$, for the stationary distributions, where $\hat{A}_1 = \hat{A}$ is the estimated population rate matrix for preseizure states (——), $\hat{A}_2 = \hat{A} + \hat{M}$ is for seizure states (——) and $\hat{V}_k$ is the estimated stationary (population) covariance matrix: (a) FP1_F7; (b) FP1_F3; (c) FP2_F4; (d) FP2_F8

5. Discussion

SDMEMs constitute an attractive class of statistical models for biomedical data. We suggested an approach for parameter inference in this framework, which even comprises more complex dynamics such as time inhomogeneity and multivariate and non-linear states. The inclusion of (deterministic) subject-specific covariate information, which causes the modeller to leave the world of identically distributed observations, is addressed as well. The conditions presented for consistency and asymptotic normality of the MLE along the lines of L_2-differentiability do not require the typical strong smoothness properties of densities and thereby open doors to irregular models. To make abstract formulations graspable, conditions are illustrated for the special case of Gaussian random effects and linear parameters (but possible non-linearity in the state). This model is a multi-dimensional extension of that studied in Delattre *et al.* (2013) and is, in its multi-dimensional version, particularly interesting as it comprises numerous well-known models. Among them are the predator–prey (or Lotka–Volterra) model (Murray, 2002), the Lorenz equations that were introduced by Lorenz (1963), which have been used to model, for example, temperature, wind speed and humidity, the Brusselator model (Kondepudi and Prigogine (2014), section 19.4), the Fitzhugh–Nagumo model (FitzHugh, 1955; Nagumo *et al.*, 1962; Jensen *et al.*, 2012), which is used to describe the regenerative firing mechanism in an excitable neuron, and the susceptible–infected–removed model that was introduced by Kermack and McKendrick (1927): an epidemic model which has been widely studied and applied (Keeling and Rohani, 2008; Guy *et al.*, 2015).

The estimation quality in terms of sample size and sampling frequency was investigated in a simulation study for a popular model in pharmacokinetics, which was motivated by a recent study (Große Ruse *et al.*, 2015). It includes subject-specific covariate information and is linear in parameters and state. When observations are sampled at high frequency, estimation results were already convincing for small sample sizes ($N = 20$), despite the comparably large number (11 fixed effects and six variances) of unknown parameters. A moderate sampling interval (of $\Delta t = 0.01$) still gave good results for the sample size considered. However, when sampling at low frequency ($\Delta t = 0.1$), the discrete time bias makes itself felt (not included here). The asymptotic normality of the MLE lends itself naturally to hypothesis tests by means of the Wald or the likelihood ratio test. On the basis of the simulated data, we estimate the false positive rate, revealing a slight liberalism of the test procedure, and compute the test's power for different true values of parameters.

Finally, we apply the framework to the statistical analysis of epileptic electroencephalography data to assess differences between dynamics for non-seizure and seizure periods. The population voltage dynamics during non-seizure and seizure states are modelled as OU processes, whereas the prevalent intersubject variability was accounted for by the inclusion of random effects in the drift. After having adjusted for the subject-specific deviations, systematic differences between preseizure and seizure recordings are assessed by comparing the population correlation structure of the corresponding stationary distributions. Our findings support those in Østergaard *et al.* (2017), which indicate increased state (channel) correlation for seizure epochs compared with non-seizure states.

A few comments are in order concerning the simulation study and the application presented. Regarding the electroencephalography data analysis, it should be noted that a physiological interpretation of our results in terms of an underlying network structure must be taken with a grain of salt for two key reasons. One is that electroencephalography recordings are only proxies for underlying brain activity. Secondly, correlation is only one way to assess signal interaction. Non-linear interactions, which are undetectable by correlation-based measures,

may still exist. In terms of our simulation settings, we have studied the method's applicability to models with up to only 17 parameters. Even in the case of an explicit likelihood, the MLE of the (unknown) covariance matrix of the random-effects vector is implicit and estimation requires numerical optimization, which may hamper estimation when the parameter space has a high dimension.

A drawback of the approach presented is the already mentioned inherent discrete time bias of the estimation procedure. It is negligible if observations are sampled at sufficiently high frequency, such as for electroencephalography recordings, but for low frequency observations, which sometimes occur in pharmacological applications, a severe bias is introduced, which should be borne in mind in applications. A possible solution is to impute data at time points in between observation times, and to conduct the estimation on the enlarged data set (Bladt *et al.*, 2016). Related to that is the problem of incomplete observations, where only some of the co-ordinates in the state space are observed, and an entire path of a completely unobserved (latent) co-ordinate should be inferred (Berg and Ditlevsen, 2013; Ditlevsen and Samson, 2014). Missing observations of one or more co-ordinates is not untypical for biological data. This, at a first step, prohibits application of the estimation procedure proposed, as it relies on the assumption of complete-data observations. Such statistical recovery of hidden state co-ordinates remains a topic for future research.

6. Asymptotic properties of the maximum likelihood estimator

In this section, we investigate asymptotic properties of the MLE and discuss some technical results.

If the drift is as in equation (2) and observations are identically distributed (in particular, the model does not contain subject-specific covariate information), consistency and asymptotic normality of the MLE can be proved by using the ideas in Delattre *et al.* (2013). The proofs are a natural extension of their setting to the multi-dimensional, affine, non-homogeneous case but become more tedious to work out in detail and to write down and will therefore be omitted here.

The classical proof of asymptotic normality of the MLE imposes strong smoothness conditions on the subject-specific density functions, such as third-order differentiability and boundedness of the derivatives. A Taylor series expansion argument together with a required asymptotic normality of the N-sample score function and a convergence of the average FI (see, for example, Bradley and Gart (1962), equation (13), or Hoadley (1971), condition N7) then yield the result. If observations are not identically distributed (e.g. if subject-specific covariate information is included in equation (1)) and the standard central limit theorem for independent and identically distribution (IID) variables cannot be applied to the score function, one can revert to the Lindeberg–Feller central limit theorem, given that the family of individual score functions $\{S^i(\theta); i \in \mathbb{N}\}$ satisfies the Lindeberg condition (a condition which limits the variation of each S^i in relation to the overall N-sample score variation). The convergence of the average FI, which is naturally given in IID models, often breaks down to requiring that covariate averages converge (Fahrmeir and Kaufmann, 1985).

The more general local asymptotic normality approach which we pursue here dispenses with the differentiability conditions by building on L_2-derivatives. More importantly, the added level of abstraction facilitates proofs, avoiding tedious and long calculations. An L_2-score function and L_2-FI are defined, which then are required to meet the above-mentioned Lindeberg and convergence conditions (see assumption (e) below and theorem 3). The first part of this section adapts results that were developed in Ibragimov and Has'minskii (2013), on consistency and

asymptotic normality of the MLE for $\theta = (\mu, \vartheta)$ in models that do not necessarily meet the differentiability conditions, to the current framework of SDMEMs with covariates. In the second part, we illustrate the verification of regularity conditions for an SDMEM with covariates and with dynamics that are frequently encountered in biomedical modelling. Although the L_2-based approach opens up for the inclusion of irregular densities in our framework, it still requires us to verify the convergence of the average FI. We shall discuss the complications of this convergence within the SDMEM framework at the end of this section.

We write $\nu^i = \mathbb{Q}^i_{\mu_0,\varphi_0}$ (see the beginning of Section 2). For simplicity, we assume that $\Theta \subseteq \mathbb{R}^q$ is open, bounded and convex and that, in all what follows, $K \subset \Theta$ is compact.

We start by stating general assumptions which the statistical model is required to satisfy and adapt them more closely to the SDMEM framework, by pointing out sufficient conditions for this particular framework which may be verified more easily. Afterwards, we establish results on asymptotic properties of the MLE for SDMEMs.

(a) $\theta \mapsto p^i(\theta)$ is ν^i almost surely continuous.

(b) $\theta \mapsto \sqrt{p^i(\theta)}$ is $L_2(\nu^i)$ differentiable with $L_2(\nu^i)$-derivative $\psi^i(\theta)$ (in other words: $p^i(\theta)$ is Hellinger differentiable). (For each θ, $\int \|\psi^i(\theta;x)\|^2 \mathrm{d}\nu^i(x) < \infty$ and $\lim_{\|h\|\to 0} \|h\|^{-2} \times \int \|\sqrt{p^i(\theta+h;x)} - \sqrt{p^i(\theta;x)} - \psi^i(\theta;x)h\|^2 \mathrm{d}\nu^i(x) = 0$.)

(c) $\psi^i(\theta)$ is continuous in $L_2(\nu^i)$. As a consequence, the matrix $I^i(\theta) = 4\int \psi^i(\theta;x)'\psi^i(\theta;x)\mathrm{d}\nu^i(x)$ exists and is continuous and will be called the FI matrix. The N-sample FI is then $I_N(\theta) = \Sigma_{i=1}^N I^i(\theta)$.

(d) The FI is bounded away from 0 and finite:

$$0 < \inf_{\theta\in\Theta} \left[\!\!\left[\frac{1}{N} I_N(\theta) \right]\!\!\right] \leqslant \sup_{\theta\in\Theta} \left[\!\!\left[\frac{1}{N} I_N(\theta) \right]\!\!\right] < \infty.$$

(e) There is a symmetric, positive definite limiting matrix $I(\theta)$ such that

$$\lim_{N\to\infty} \sup_{\theta\in K} \left[\!\!\left[\frac{1}{N} I_N(\theta) - I(\theta) \right]\!\!\right] = 0$$

and

$$\lim_{N\to\infty} \sup_{\theta\in K} \left[\!\!\left[\left\{ \frac{1}{N} I_N(\theta) \right\}^{-1/2} - I(\theta)^{-1/2} \right]\!\!\right] = 0.$$

Analogously to the traditional setting, we call $S^i(\theta) = 2p^i(\theta)^{-1/2}\psi^i(\theta)$ the score function of sample i and set $S_N(\theta) = \Sigma_{i=1}^N S^i(\theta)$ for the N-sample score function. One can show that also in this more general setting the score function is centred (Ibragimov and Has'minskii (2013), page 115).

Sufficient conditions for the almost sure continuity of $p^i(\theta)$ in θ are continuity of $\mu \mapsto q^i(\mu,\varphi)$ and of $\vartheta \mapsto g(\varphi;\vartheta)$, together with the existence of an integrable function of φ dominating $q^i(\mu,\varphi)g(\varphi;\vartheta)$. Continuity of g holds for instance in the common case where g is a Gaussian density $\mathcal{N}(0,\vartheta)$ and ϑ is bounded away from 0. For conditions on the continuity of q^i, suppose that F is continuous and assume for simplicity that $\Sigma(t,x) \equiv I$ is the identity matrix. If $\mu \mapsto F(X^i_s, D^i_s, \mu, \varphi)$ is uniformly continuous (for instance differentiable with bounded Jacobian), then $\mu \mapsto \int_0^{T^i} F(X^i_s, D^i_s, \mu, \varphi)' F(X^i_s, D^i_s, \mu, \varphi)\mathrm{d}s$ is continuous. If F moreover has the property that, for some $\kappa, C > 0$, $\|F(X^i, D^i_s, \mu, \varphi) - F(X^i, D^i_s, \mu_0, \varphi)\| \leqslant C(1 + \|X^i\|^\kappa)\|\mu - \mu_0\|$, Kolmogorov's continuity criterion guarantees continuity of q^i.

The L_2-differentiability is neither stronger nor weaker than standard (pointwise) differentiability. Generally, none implies the other but, under certain conditions, the limits are identical. Of course, if p^i is L_2 differentiable and differentiable in the ordinary sense, then

$$\psi^i(\theta;x) = \frac{\mathrm{d}}{\mathrm{d}\theta} p^i(\theta;x)^{1/2}.$$

To point out the connection between the FI and score functions defined via L_2-derivatives and their counterparts based on 'standard' differentiability, we recall the following result (van der Vaart (2000), lemma 7.6). If $\theta \mapsto \sqrt{p^i(\theta)}$ is continuously differentiable, the quantity

$$\tilde{S}^i(\theta) := 2p^i(\theta)^{-1/2} \frac{\mathrm{d}}{\mathrm{d}\theta} p^i(\theta)$$

is well defined (since $p^i > 0$). If $\tilde{I}^i(\theta) = \mathbb{E}_\theta\{\tilde{S}^i(\theta)\tilde{S}^i(\theta)'\}$ is finite and, continuous, $\theta \mapsto \sqrt{p^i(\theta)}$ is L_2 differentiable, the L_2-derivative coincides with the pointwise derivative and, in fact, $\tilde{S}^i(\theta) = S^i(\theta)$ and $\tilde{I}^i(\theta) = I^i(\theta)$.

Note as well that the assumption on the (norm of the) FI matrix to grow beyond bounds (see condition (e)) corresponds to the requirement of infinite flow of information. This is naturally connected to the consistency of estimators.

In what follows, we write briefly and somewhat sloppily θ_N if it is of the form $\theta_N = \theta + I_N(\theta)^{-1/2}h$ for some $\theta \in K$ and $h \in \Theta_{N,\theta} = \{h \in \mathbb{R}^q : \theta + I_N(\theta)^{-1/2}h \in \Theta\}$.

We can now state results on the asymptotic behaviour of the MLE in SDMEMs with covariates. These are consequences of theorems in Ibragimov and Has'minskii (2013), and proofs are only briefly outlined.

Theorem 2 (consistency). The MLE of model (1) is uniformly on K consistent, if (condition 1) there is a constant $m > q$ such that $\sup_{\theta\in K} \mathbb{E}_\theta\{\|S_N(\theta)\|^m\} < \infty$, and (condition 2) there is a positive constant $a(K)$ such that for (sufficiently large N and) all $\theta \in K$ (and all $h \in \Theta_{N,\theta}$)

$$H_i^2(\theta, \theta_N) \geqslant a(K) \frac{\|\theta_N - \theta\|^2}{1 + \|\theta_N - \theta\|^2},$$

where $H_i^2(\theta_1, \theta_2) := \int \{\sqrt{p^i(\theta_1)} - \sqrt{p^i(\theta_2)}\}^2 \mathrm{d}\nu^i$ is the squared Hellinger distance between $\mathbb{Q}^i_{\theta_1}$ and $\mathbb{Q}^i_{\theta_2}$.

Proof. Condition 1 is an extension of lemma III.3.2 in Ibragimov and Has'minskii (2013) to non-homogeneous observations. Condition 2 is adapted from Ibragimov and Has'minskii (2013), lemma I.5.3.

Remark 2. If the dimension of the parameter set is 1, condition 1 can be replaced by a subquadratic growth condition on the Hellinger distance (for IID observations, see Ibragimov and Has'minskii (2013), theorem I.5.3), namely that $H^2(\theta_1, \theta_2) \leqslant A\|\theta_2 - \theta_1\|^2$, such that consistency here reduces to $H^2(\theta_1, \theta_2)$ behaving asymptotically as $\|\theta_2 - \theta_1\|^2$.

The following theorem establishes the so-called uniform asymptotic normality of the model, which in turn implies the asymptotic normality of the MLE (theorems II.6.2 and III.1.1, Ibragimov and Has'minskii (2013)).

Theorem 3 (asymptotic normality). Assume conditions 1 and 2 from theorem 2 and additionally (condition 3) the family $\{S^i(\theta), i = 1, \ldots, N\}$ satisfies the Lyapunov condition uniformly in K, i.e. there is $\delta > 0$ such that

$$\lim_{N\to\infty} \sup_{\theta\in K} \sum_{i=1}^{N} \mathbb{E}_\theta\{\|I_N(\theta)^{-1/2} S^i(\theta)\|^{2+\delta}\} = 0,$$

and (condition 4) $\forall\ R > 0$,

$$\lim_{N\to\infty}\sup_{\theta\in K}\sup_{\|h\|<R}\sum_{i=1}^{N}\int[\{\psi^i(\theta_N)-\psi^i(\theta)\}I_N(\theta)^{-1/2}h]^2\mathrm{d}\nu^i=0.$$

Then $\{\hat{\theta}_N\}_{N\in\mathbb{N}}$ is uniformly in K consistent and asymptotically Gaussian distributed with parameters $(\theta, I_N(\theta)^{-1})$, and all moments of $\{I_N(\theta)^{1/2}(\hat{\theta}_N-\theta)\}_{N\in\mathbb{N}}$ converge uniformly in K to the corresponding moments of the $\mathcal{N}(0, I)$ distribution.

Condition (3) can be generalized to the Lindeberg condition. If the densities $\sqrt{p^i(\theta)}$ are twice continuously differentiable with second derivative $J^i(\theta)$, assumption 4 can be replaced by requiring that

$$\lim_{N\to\infty}\sup_{\theta\in K}\sup_{\|h\|\leqslant R}[\![I_N(\theta)^{-1/2}]\!]^4\sum_{i=1}^{N}\int[\![J^i(\theta_N)]\!]^2\mathrm{d}\nu^i=0.$$

As pointed out in Section 1, for a general SDMEM the p^i are not explicitly available. One can, however, formulate conditions for the drift function F and the random-effects density g, which implicitly guarantee the differentiability of $\log\{p^i(\theta)\}=\log\{\int q^i(\mu,\varphi)g(\varphi;\vartheta)\mathrm{d}\varphi\}$. This can, for example, be done by assuring that differentiation can be passed under the integral sign: sufficient conditions for the differentiability of $\log\{p^i(\theta)\}$ with respect to μ would, for example, include differentiability of q^i with respect to μ and a uniform-in-μ domination of

$$\frac{\mathrm{d}}{\mathrm{d}\mu}q^i(\mu,\varphi)\left\{\int q^i(\mu,\varphi)g(\varphi;\vartheta)\mathrm{d}\varphi\right\}^{-1}.$$

However, explicitly formulating these conditions for a generic SDMEM is not illustrative; suitable conditions should be formulated and checked for the specific application at hand. One particular case in which the $p^i(\theta)$ are explicitly available is the affine model (2), which we consider in more detail below.

6.1. Stochastic differential mixed effect model with covariates and affine mixed effects

We illustrate the verification of the assumptions for theorems 2 and 3 for the affine SDMEM (2). The estimation performance and hypothesis testing for various sample sizes and sampling frequencies of this model were studied in Section 3, and applied in Section 4 for the statistical investigation of electroencephalography data. For simplicity we assume that $B=C$, such that $U_i:=U_{1i}=U_{2i}$ and $V_i=V_{1i}=V_{2i}=Z_i$. The likelihood (4) can be written as

$$p^i(\theta)=\frac{1}{\sqrt{\det(I+V_i\Omega)}}\exp\left\{-\frac{1}{2}(\mu-V_i^{-1}U_i)'G^i(\Omega)(\mu-V_i^{-1}U_i)\right\}\exp\left(\frac{1}{2}U_i'V_i^{-1}U_i\right),$$

with $G^i(\Omega)=(I+V_i\Omega)^{-1}V_i$. Defining $\gamma^i(\theta)=G^i(\Omega)(V_i^{-1}U_i-\mu)$ (we assume that V_i is almost surely invertible), the score function for subject i is thus given by

$$S^i(\theta)=\left[\frac{\mathrm{d}}{\mathrm{d}\mu}\log\{p^i(\theta)\},\frac{\mathrm{d}}{\mathrm{d}\Omega}\log\{p^i(\Omega)'\}\right],$$

with

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\log\{p^i(\theta)\}=\gamma^i(\theta)'$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\Omega}\log\{p^i(\theta)\}=\tfrac{1}{2}\{-G^i(\Omega)+\gamma^i(\theta)\gamma^i(\theta)'\}.$$

We start by verifying condition 1. Since the set $K \subset \Theta$ is compact, there are positive constants A_K, B_K and C_K such that $\|\mu\| \leqslant A_K, B_K \leqslant [\![\Omega]\!] \leqslant C_K$. One can show that $[\![G^i(\Omega)]\!] \leqslant [\![\Omega^{-1}]\!]$, which gives the upper bound $\|S^i(\theta)\| \leqslant (\|\gamma^i(\theta)\| + [\![\Omega^{-1}]\!] + \|\gamma^i(\theta)\|^2)$. Moreover, the moment-generating function $\Phi_{\theta,\gamma^i(\theta)}(a)$ of $\gamma^i(\theta)$ can be bounded from above by $\exp(\frac{1}{2}a'\Omega' a)$, for $a \in \mathbb{R}^d$. This can be used to find that $\mathbb{E}_\theta\{\|\gamma_i(\theta)\|^m\} \leqslant C_1$ for some constant C_1 that may depend on K, d and m. Therefore, there is another constant C_2, which may depend on K, d, m and N, such that $\mathbb{E}_\theta\{\|S_N(\theta)^m\} \leqslant C_2$, proving condition 1.

To verify condition 2, note that the regularity of $p_N(\theta) = \Pi_{i=1}^N p^i(\theta)$ and its derivatives implies that

$$\begin{aligned} H^2(\theta, \theta_N) &= \int \{-\psi_N(\theta)(\theta_N - \theta) + \sqrt{p_N(\theta_N)} - \sqrt{p_N(\theta)} + \psi_N(\theta)(\theta_N - \theta)\}^2 \mathrm{d}\nu \\ &= \int \{-\psi_N(\theta)(\theta_N - \theta)\}^2 \mathrm{d}\nu + o(\|\theta_N - \theta^2) \\ &= (\theta_N - \theta)' I_N(\theta)(\theta_N - \theta) + o(\|\theta_N - \theta\|^2) - 2O(\|\theta_N - \theta\|^2) o(\|\theta_N - \theta\|^2) \\ &\geqslant \|(\theta_N - \theta)\|^2 \lambda_{N,\min}(\theta) + o(\|\theta_N - \theta\|^2), \end{aligned}$$

where $\lambda_{N,\min}(\theta)$ denotes the smallest eigenvalue of $I_N(\theta)$. Therefore, for N sufficiently large, there is a constant A_K such that $H^2(\theta, \theta_N) \geqslant A_K \|(\theta_N - \theta)\|^2$. Since Θ is bounded, we even have

$$\|(\theta_N - \theta)\|^2 \geqslant C \frac{\|(\theta_N - \theta)\|^2}{1 + \|(\theta_N - \theta)\|^2}$$

for some positive constant C, which shows that condition 2 holds.

The Lyapunov condition 3 follows in a straightforward way. According to above, $\mathbb{E}_\theta\{\|S^i(\theta)\|^3\} \leqslant C$ for some C and therefore

$$\begin{aligned} \sup_{\theta \in K} \sum_{i=1}^N \mathbb{E}_\theta\{\|I_N(\theta)^{-1/2} S^i(\theta)^3\} &\leqslant N^{-3/2} \sup_{\theta \in K} [\![\sqrt{N} I_N(\theta)^{-1/2} - I(\theta)^{-1/2}]\!] \sum_{i=1}^N \mathbb{E}_\theta\{\|S^i(\theta)\|^3\} \\ &\quad + N^{-3/2} \sup_{\theta \in K} [\![I(\theta)^{-1/2}]\!] \sum_{i=1}^N \mathbb{E}_\theta\{\|S^i(\theta)\|^3\} \\ &\leqslant C N^{-1/2} \left\{ \sup_{\theta \in K} [\![\sqrt{N} I_N(\theta)^{-1/2} - I(\theta)^{-1/2}]\!] + \sup_{\theta \in K} [\![I(\theta)^{-1/2}]\!] \right\}, \end{aligned}$$

which converges to 0 as $N \to \infty$.

To verify condition 4, we show that (recall that $J^i(\theta)$ denotes the second derivative of $\sqrt{p^i(\theta)}$)

$$\begin{aligned} &\sup_{\|h\| \leqslant R} \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\nu^i} \left\{ [\![J^i(\theta_N) - J^i(\theta)]\!]^2 \right\} \right], \\ &\qquad \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\nu^i} \left\{ [\![J^i(\theta)]\!]^2 \right\} \right] \end{aligned} \tag{8}$$

converge to 0 uniformly in K. As $J^i(\theta)$ is continuous, it is uniformly continuous on compact sets, such that, for all $i \in \mathbb{N}$, $a_{i,N} = \sup_{\|h\| \leqslant R} [\![J^i(\theta_N) - J^i(\theta)]\!]$ converges almost surely to 0 as $N \to \infty$. One can show that $a_{i,N} \leqslant A^i(\theta, R)$ and $\mathbb{E}_{\nu^i}\{A^i(\theta, R)^2\} \leqslant D_K$. Dominated convergence implies that $\mathbb{E}_\theta(a_{i,N}) \to 0$, and the uniform (in i) bound D_K implies uniform in K convergence of the first term in expression (8) to 0. For the second term in expression (8) we note that

$$\mathbb{E}_{\nu^i}\{[\![J^i(\theta)]\!]^2\} \leqslant \mathbb{E}_\theta\left\{\left[\!\!\left[\frac{\mathrm{d}}{\mathrm{d}\theta}S^i(\theta)\right]\!\!\right]^2\right\} + \mathbb{E}_\theta\{[\![S^i(\theta)'S^i(\theta)]\!]^2\} < C_K,$$

where C_K is a constant that depends on K only and we conclude uniform in $\theta \in K$ convergence to 0, completing the verification of condition 4.

6.2. On the convergence of the average Fisher information in stochastic differential mixed effects models

As seen above, a key condition for establishing the asymptotic normality of the MLEs was the convergence of the scaled N-sample FI $(1/N)I_N(\theta) = (1/N)\Sigma_{i=1}^N I^i(\theta)$ to a deterministic limit $I(\theta)$. This is difficult to check when the drift contains subject-specific covariate information D^i and these covariates are not identical across subjects, because the processes X^i do not have the same distributions, since the drift function F varies across subjects, $F^i(x, \mu, \phi^i) = F(x, D^i, \mu, \phi^i)$.

In a linear regression model with random effects, the asymptotic behaviour of the averaged FI is deduced from a comparable asymptotic behaviour of the averaged covariates, such that the verification of the conditions can conveniently be accomplished on the covariate level. Also, in SDMEMs with covariates, it would be desirable to be able to break down the convergence of $(1/N)I_N(\theta)$ to an average covariate behaviour. This, however, is not possible, not even if we assume the simplest case where the drift function F is linear in state, covariates and fixed and random effects and if the random effects are Gaussian distributed with known covariance matrix.

We illustrate this in the simplest non-trivial example that includes covariates. We look at a one-dimensional state process X^i governed by $\mathrm{d}X_t^i = \{X_t^i(\mu^1 + \phi^{i,1}) + D_t^i(\mu^2 + \phi^{i,2})\}\mathrm{d}t + \mathrm{d}W_t^i$, with fixed effects vector $\mu = (\mu^1, \mu^2)'$, IID $\mathcal{N}(0, \Omega)$-distributed random effects $\phi^i = (\phi^{i,1}, \phi^{i,2})'$ and known covariate process D^i. We assume that Ω is known, such that $\theta = \mu$ is the only unknown parameter. This set-up is a special case of equation (2) with $A = 0$ and $B = C$ and therefore $U_i := U_{1i} = U_{2i}$ and $V_i := V_{1i} = V_{2i} = Z_i$. More specifically,

$$U_i = \begin{pmatrix} \int_0^T X_t^i \mathrm{d}X_t^i \\ \int_0^T D_t^i \mathrm{d}X_t^i \end{pmatrix}$$

and

$$V_i = \begin{pmatrix} \int_0^{T^i} (X_t^i)^2 \mathrm{d}t & \int_0^{T^i} X_t^i D_t^i \mathrm{d}t \\ \int_0^{T^i} X_t^i D_t^i \mathrm{d}t & \int_0^{T^i} (D_t^i)^2 \mathrm{d}t \end{pmatrix}.$$

The FI is by definition

$$I^i(\mu) = \mathbb{E}_\mu\left[-\frac{\mathrm{d}^2}{\mathrm{d}\mu^2}\log\{p^i(\theta)\}\right] = \mathbb{E}_\mu\{(I + V_i\Omega)^{-1}V_i\};$$

see equation (4). The matrix $(I + V_i\Omega)V_i$ is, however, a non-linear function of V_i and thus finding an explicit expression for $I^i(\mu)$ is generally impossible—even in the simple linear case, where X^i is nothing but a Gaussian process. For comparison, in the linear mixed effects

model, the log-likelihood for observation y^i with covariate vectors x^i and z^i is proportional to $-\frac{1}{2}(y^i-(x^i)'\mu)'(I+z_i'\Omega z_i)^{-1}(y^i-(x^i)'\mu)$, and therefore the FI is $\mathbb{E}_\mu\{x^i(I+z_i'\Omega z_i)^{-1}(x^i)'\}=x^i(I+z_i'\Omega z_i)^{-1}(x^i)'$. The crucial difference, compared with the linear SDMEM case is that the matrix $(I+z_i'\Omega z_i)^{-1}$ is deterministic. Therefore, convergence of $(1/N)\Sigma_{i=1}^N I^i(\theta)$ is implied by a limiting behaviour of averages. This is particularly attractive as one can often design the experiment in such a way that the required limiting behaviour holds. In the SDMEM case, however, it will generally not be possible to determine from an analytical expression of $I_N(\theta)$ whether the condition $(1/N)I_N(\theta)\to I(\theta)$ holds, because of the combination of non-linearity and stochasticity.

Acknowledgements

The work is part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen. Adeline Samson has been partially supported by the LabExPERSYVAL-Lab (ANR-11-LABX-0025-01).

References

Aït-Sahalia, Y. (2002) Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, **70**, 223–262.

Beal, S. L. and Sheiner, L. B. (1981) Estimating population kinetics. *Crit. Rev. Biomed. Engng*, **8**, no. 3, 195–222.

Berg, R. W. and Ditlevsen, S. (2013) Synaptic inhibition and excitation estimated via the time constant of membrane potential fluctuations. *J. Neurphysiol.*, **110**, 1021–1034.

Bladt, M., Finch, S. and Sørensen, M. (2016) Simulation of multivariate diffusion bridges. *J. R. Statist. Soc.* B, **78**, 343–369.

Bradley, R. A. and Gart, J. J. (1962) The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika*, **49**, 205–214.

Delattre, M., Genon-Catalot, V. and Laredo, C. (2018) Parametric inference for discrete observations of diffusion processes with mixed effects. *Stoch. Processes Appl.*, **128**, 1929–1957.

Delattre, M., Genon-Catalot, V. and Samson, A. (2013) Maximum likelihood estimation for stochastic differential equations with random effects. *Scand. J. Statist.*, **40**, 322–343.

Delattre, M., Genon-Catalot, V. and Samson, A. (2015) Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. *Eur. Ser. Appl. Indust. Math. Probab. Statist.*, **19**, 671–688.

Delattre, M., Genon-Catalot, V. and Samson, A. (2016) Mixtures of stochastic differential equations with random effects: application to data clustering. *J. Statist. Plannng Inf.*, **173**, 109–124.

Delyon, B., Lavielle, V. and Moulines, E. (1999) Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, **27**, 94–128.

Ditlevsen, S. and Samson, A. (2014) Estimation in the partially observed stochastic Morris–Lecar neuronal model with particle filter and stochastic approximation methods. *Ann. Appl. Statist.*, **8**, 674–702.

Donnet, S., Foulley, J.-L. and Samson, A. (2010) Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics*, **66**, 733–741.

Durham, G. B. and Gallant, A. R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econ. Statist.*, **20**, 297–338.

Fahrmeir, L. and Kaufmann, H. (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.*, **13**, 342–368.

FitzHugh, R. (1955) Mathematical models of threshold phenomena in the nerve membrane. *Bull. Math. Biophys.*, **17**, 257–278.

Große Ruse, M., Søndergaard, L. R., Ditlevsen, S., Damgaard, M., Fuglsang, S., Ottesen, J. T. and Madsen, J. L. (2015) Absorption and initial metabolism of 75 se-l-selenomethionine: a kinetic model based on dynamic scintigraphic data. *Br. J. Nutrn*, **114**, 1718–1723.

Guy, R., Larédo, C. and Vergu, E. (2015) Approximation of epidemic models by diffusion processes and their statistical inference. *J. Math. Biol.*, **70**, 621–646.

Hoadley, B. (1971) Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Ann. Math. Statist.*, **42**, 1977–1991.

Ibragimov, I. A. and Has'minskii, R. Z. (2013) *Statistical Estimation: Asymptotic Theory*. New York: Springer.

Jensen, A. C., Ditlevsen, S., Kessler, M. and Papaspiliopoulos, O. (2012) Markov chain Monte Carlo approach to parameter estimation in the Fitzhugh-Nagumo model. *Phys. Rev.* E, **86**, no. 4, article 041114.

Keeling, M. J. and Rohani, P. (2008) *Modeling Infectious Diseases in Humans and Animals*. Princeton: Princeton University Press.
Kermack, W. O. and McKendrick, A. G. (1927) A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond.* A, **115**, 700–721.
Kessler, M., Lindner, A. and Sørensen, M. (2012) *Statistical Methods for Stochastic Differential Equations*. Boca Raton: CRC Press.
Kondepudi, D. and Prigogine, I. (2014) *Modern Thermodynamics: from Heat Engines to Dissipative Structures*. Chichester: Wiley.
Leander, J., Lundh, T. and Jirstrand, M. (2014) Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. *Math. Biosci.*, **251**, 54–62.
Le Cam, L. (2012) *Asymptotic Methods in Statistical Decision Theory*. New York: Springer.
Lehmann, E. L. and Romano, J. P. (2006) *Testing Statistical Hypotheses*. New York: Springer Science and Business Media.
Lo, A. W. (1988) Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econmetr. Theory*, **4**, 231–247.
Lorenz, E. N. (1963) Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
Maitra, T. and Bhattacharya, S. (2018a) On classical and Bayesian asymptotics in stochastic differential equations with random effects having mixture normal distributions. *Preprint arXiv:1605.03333v3*. Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, Baranagar.
Maitra, T. and Bhattacharya, S. (2018b) On asymptotic inference in stochastic differential equations with time-varying covariates. *Can. J. Statist.*, **46**, 635–655.
Møller, J. B., Overgaard, R. V., Madsen, H., Hansen, T., Pedersen, O. and Ingwersen, S. H. (2010) Predictive performance for population models using stochastic differential equations applied on data from an oral glucose tolerance test. *J. Pharmkinet. Pharmdyn.*, **37**, 85–98.
Murray, J. D. (2002) *Mathematical Biology I: an Introduction*. New York: Springer.
Nagumo, J., Arimoto, S. and Yoshizawa, S. (1962) Active pulse transmission line simulating nerve axon. *Proc. Inst. Radio Engrs*, **50**, 2061–2070.
Østergaard, J., Rahbek, A. and Ditlevsen, S. (2017) Oscillating systems with cointegrated phase processes. *J. Math. Biol.*, **75**, 845–883.
Pedersen, A. R. (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, **22**, 55–71.
Phillips, P. C. and Yu, J. (2009) Maximum likelihood and Gaussian estimation of continuous time models in finance. In *Handbook of Financial Time Series* (eds T. Mikosch, J. P. Kreiss, R. A. Davis and T. G. Andersen), pp. 497–530. New York: Springer.
Picchini, U., De Gaetano, A. and Ditlevsen, S. (2010) Stochastic differential mixed-effects models. *Scand. J. Statist.*, **37**, 67–90.
Picchini, U. and Ditlevsen, S. (2011) Practical estimation of high dimensional stochastic differential mixed-effects models. *Computnl Statist. Data Anal.*, **55**, 1426–1444.
Picchini, U., Ditlevsen, S., De Gaetano, A. and Lansky, P. (2008) Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. *Neurl Computn*, **20**, 2696–2714.
Picchini, U. and Forman, J. L. (2019) Bayesian inference for stochastic differential equation mixed effects models of a tumour xenography study. *Appl. Statist.*, **68**, 887–913.
Prado, R., West, M. and Krystal, A. D. (2001) Multichannel electroencephalographic analyses via dynamic regression models with time-varying lag–lead structure. *Appl. Statist.*, **50**, 95–109.
Qin, L. and Wang, Y. (2008) Nonparametric spectral analysis with applications to seizure characterization using EEG time series. *Ann. Appl. Statist.*, **2**, 1432–1451.
Rügamer, D., Brockhaus, S., Gentsch, K., Scherer, K. and Greven, S. (2018) Boosting factor-specific functional historical models for the detection of synchronization in bioelectrical signals. *Appl. Statist.*, **67**, 621–642.
Shoeb, A. H. (2009) Application of machine learning to epileptic seizure onset detection and treatment. *PhD Thesis*. Massachusetts Institute of Technology, Cambridge.
Stephen, E. P., Lepage, K. Q., Eden, U. T., Brunner, P., Schalk, G., Brumberg, J. S., Guenther, F. H. and Kramer, M. A. (2014) Assessing dynamics, spatial scale, and uncertainty in task-related brain network analyses. *Front. Computnl Neursci.*, **8**, article 31.
van der Vaart, A. W. (2000) *Asymptotic Statistics*, vol. 3. Cambridge: Cambridge University Press.
Wolfinger, R. (1993) Laplace's approximation for nonlinear mixed models. *Biometrika*, **80**, 791–795.
Wulsin, D. F., Jensen, S. T. and Litt, B. (2016) Nonparametric multi-level clustering of human epilepsy seizures. *Ann. Appl. Statist.*, **10**, 667–689.
Zhang, T., Wu, J., Li, F., Caffo, B. and Boatman-Reich, D. (2015) A dynamic directional model for effective brain connectivity using electrocorticographic (ECoG) time series. *J. Am. Statist. Ass.*, **110**, 93–106.