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

Mareile Große Ruse

Department of Mathematical Sciences, University of Copenhagen.
mareile@math.ku.dk

Adeline Samson

Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France.
adeline.leclercq-samson@imag.fr

Susanne Ditlevsen

Department of Mathematical Sciences, University of Copenhagen.
susanne@math.ku.dk

Summary. Neurobiological data such as EEG 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 modeling framework for this type of data based on stochastic processes. Stochastic differential equations with mixed effects are a popular framework for modeling biomedical data, e.g., in pharmacological studies. While the inherent stochasticity of diffusion models accounts for prevalent model uncertainty or misspecification, random effects take care of the inter-subject variability. The 2-layer stochasticity, however, renders parameter inference challenging. This is especially true for more complex model dynamics, and only few theoretical investigations on the asymptotic behavior of estimates exist. This article adds to filling this gap by examining asymptotics (number of subjects going to infinity) of Maximum Likelihood estimators in multidimensional, nonlinear and non-homogeneous stochastic differential equations with random effects and included covariates. Estimates are based on the discretized continuous-time likelihood and we investigate finite-sample and discretization bias. In applications, the comparison of, e.g., different experimental conditions such as *placebo* vs. *treatment*, is often of interest. We suggest a hypothesis testing approach and evaluate the test's performance by simulations. Finally, we apply the framework to a statistical investigation of EEG recordings from epileptic patients.

1. Introduction

Many biomedical studies are based on image data, which is characterized by a high time resolution, but also a low signal-to-noise ratio. The same happens with EEG data, which are measurements of electrical activitiy 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 modeled by continous 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, for example non-specified feedback mechanisms or genetic variation. The intra-subject variability in longitudinal data asks for a model that incorporates system noise. Any systematic inter-subject variability is usually well explained by the inclusion of covariate information, e.g., treatment regime, gender or specifics of experimental conditions. The remaining inter-subject variability can then be taken care of by random effects. The motivating examples for our work are EEG measurements from multiple channels, and 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 to the typical time scales of the observed system. This allows us to employ techniques facilitating the use of continuous-time stochastic processes. We therefore propose a new modeling 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 inter-subject variability through covariates and random effects.

Models that combine SDEs and random effects (i.e., so-called stochastic differential mixed-effects models, SDMEMs) have become a popular framework for modeling biological data (Guy et al., 2015; Donnet et al., 2010; Møller et al., 2010; Leander et al., 2014; Picchini et al., 2008). They come with three advantages: Firstly, they capture inter-subject 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. The latter is due to the fact that the mixed-effects approach facilitates pooling of data across subjects, which leads to unbiasedness of the drift estimator as the number of subjects approaches infinity.

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 (nonlinear) SDE models is analytically not available, rendering parameter inference for standard SDE models a nontrivial problem in itself. On the other hand, the likelihood has to 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 latter 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 is 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 nonlinear mixed-effects models are first-order conditional estimation (FOCE) (Beal and Sheiner, 1981) and Laplace approximation (Wolfinger, 1993). A computational alternative is the expectation-maximization (EM) 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 modeled 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 are investigated in Delattre et al. (2015, 2017). The case of continuous-time observations of a univariate SDMEM with Gaussian mixed effects entering the drift linearly is considered in Delattre et al. (2013).

Two aspects that are important in modeling biomedical data are not covered by these works: On the one hand, the theoretical investigations of estimators when the state process is modeled by a multivariate, time-inhomogeneous and nonlinear SDE, and on the other hand, the inclusion of covariate information. The lack of both in a model implies considerable restrictions for practitioners and the purpose of this article is to fill this gap.

If the drift function is linear in the parameters, the standard asymptotic properties of the MLE in multidimensional, time-homogeneous, nonlinear 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 (multidimensionality of the state, nonlinearity, 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 setup. Their approach has, however, two drawbacks. The first one 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 imposed regularity assumptions are rather restrictive, for instance, the density of the random effects may not be smooth. If, for instance, random effects are supposed to have a double exponential distribution, those regularity conditions are not met. However, the Laplace density, e.g., is "almost" regular, satisfying a particular type of first-order differentiability. This almost regularity can be treated by the more general approach which builds upon L_2-differentiability and the local asymptotic normality (LAN) 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 LAN perspective.

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

The article is structured as follows. Section 2 introduces the model framework, investigates asymptotic properties of the MLE and applies the asymptotic results to hypothesis testing. Moreover, we exemplify the framework with covariates for affine mixed effects. Section 3 is devoted to a simulation study in a model 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 EEG recordings of epileptic patients in section 4, with the purpose of investigating how channel interactions differ between non-seizure and seizure states. Finally, we conclude with a discussion.

2.   Maximum likelihood estimation for SDMEMs 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 $X^i = (X_t^i)_{0 \leq t \leq T^i}$ whose dynamics are governed by the stochastic differential equations

$$dX_t^i = F(X_t^i, D_t^i, \mu, \phi^i)dt + \Sigma(t, X_t^i)dW_t^i, \quad 0 \leq t \leq T^i, \quad X_0^i = x_0^i, \quad i = 1, \ldots, N. \quad (1)$$

The r-dimensional Wiener processes $W^i = (W_t^i)_{t \geq 0}$ and the d-dimensional random vectors ϕ^i are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which is rich enough to ensure independence of all random objects $W^i, \phi^i, i = 1, \ldots, N$. The d-dimensional vectors $\phi^i, i = 1, \ldots, N$, are the so-called *random effects*. They are assumed to be $\mathcal{F}_0$-measurable and have a common (usually centered) distribution which is specified by a (parametrized) Lebesgue density $g(\varphi; \vartheta)d\varphi$. 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 dependency, 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 $T = \max_{1 \leq i \leq N} T^i$, 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 (i) existence and uniqueness of the solution to (1) and (ii) existence and good behaviour of the Radon-Nikodym derivative

$$q^i(\mu, \varphi) := q^i(\mu, \varphi; X^i) = \frac{d\mathbb{Q}_{\mu,\varphi}^i}{d\mathbb{Q}_{\mu_0,\varphi_0}^i}(X^i)$$

$$= \exp\left(\int_0^{T^i} \left[F(X_s^i, D_s^i, \mu, \varphi) - F(X_s^i, D_s^i, \mu_0, \varphi_0) \right]' \Gamma^{-1}(s, X_s^i) dX_s^i \right.$$

$$\left. - \frac{1}{2} \int_0^{T^i} \left[F(X_s^i, D_s^i, \mu, \varphi) - F(X_s^i, D_s^i, \mu_0, \varphi_0) \right]' \Gamma^{-1}(s, X_s^i) \left[F(X_s^i, D_s^i, \mu, \varphi) + F(X_s^i, D_s^i, \mu_0, \varphi_0) \right] ds \right),$$

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

We observe X^i at time points $0 \leq t_0^i < t_1^i < \ldots < t_{n_i}^i = T^i$ and the inference task consists in recovering the "true" underlying θ based on observations $X_{t_0^i}^i, \ldots, X_{t_n^i}^i, i = 1, \ldots, N$. We approach this inference task by first supposing to have the entire paths $(X_t^i)_{0 \leq t \leq T^i}, i = 1, \ldots, N$, at our disposal. Based on these we derive the continuous-time MLE and discretize it in a second step. The bias 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,

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

An example of (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 analyze EEG 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 (2) enables the modeler to impose random effects on only a selection of fixed effects. The conditional likelihood turns into the compact expression $q^i(\mu, \varphi) = \mathrm{e}^{\mu' U_{1i} - \frac{1}{2}\mu' V_{1i}\mu + \varphi' U_{2i} - \frac{1}{2}\varphi' V_{2i}\varphi - \varphi' Z_i \mu}$ with the sufficient statistics

$$
\begin{aligned}
U_{1i} &= \int_0^{T^i} B(X_s^i, D_s^i)' \Gamma^{-1}(s, X_s^i) \left[dX_s^i - A(X_s^i, D_s^i)ds \right], \\
V_{1i} &= \int_0^{T^i} B(X_s^i, D_s^i)' \Gamma^{-1}(s, X_s^i) B(X_s^i, D_s^i)ds, \\
U_{2i} &= \int_0^{T^i} C(X_s^i, D_s^i)' \Gamma^{-1}(s, X_s^i) \left[dX_s^i - A(X_s^i, D_s^i)ds \right], \\
V_{2i} &= \int_0^{T^i} C(X_s^i, D_s^i)' \Gamma^{-1}(s, X_s^i) C(X_s^i, D_s^i)ds, \\
Z_i &= \int_0^{T^i} C(X_s^i, D_s^i)' \Gamma^{-1}(s, X_s^i) B(X_s^i, D_s^i)ds.
\end{aligned}
\tag{3}
$$

Integration over φ gives the unconditional likelihood for subject i,

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

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

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

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

Discrete data

Above we assumed to observe the entire paths $(X_t^i)_{0 \leq t \leq T}$. In practice, observations are only available at discrete time points $t_0, \ldots, t_n$. 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 $\int_{t_k}^{t_{k+1}} h(s, X_s^i) dX_s^i$ 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 $t_k = T\frac{k}{n}$ and denote by $U_{1i}^n, V_{1i}^n, U_{2i}^n, V_{2i}^n, Z_i^n$ the first-order discrete-time approximations to the continuous-time statistics $U_{1i}, V_{1i}, U_{2i}, V_{2i}, Z_i$ in eq. (3), one has the following result:

THEOREM 1 (NEGLIGIBILITY OF DISCRETIZATION ERROR).
Assume model (2) *and suppose that* $A, B'\Gamma^{-1}B, B'\Gamma^{-1}C, C'\Gamma^{-1}C, B'\Gamma^{-1}, C'\Gamma^{-1}$ *are globally Lipschitz-continuous in* (t, x) *and that in addition to* A, B, C *and* Σ *also* $B'\Gamma^{-1}, C'\Gamma^{-1}$ *is of sublinear growth in* x, *uniformly in* t. *Then, for all* $p \geq 1$ *and all* $i = 1, \ldots, N$, *there is a constant* K *such that*

$$\mathbb{E}_{\theta_0} \left([\![V_{1i} - V_{1i}^n]\!]^p + \|U_{1i} - U_{1i}^n\|^p + [\![V_{2i} - V_{2i}^n]\!]^p + \|U_{2i} - U_{2i}^n\|^p + \|Z_i - Z_i^n\|^p \right) \leq K \left(\frac{T}{n} \right)^{p/2}.$$

The discretization error is investigated numerically in section 3.

2.3. Asymptotic properties of the MLE

If the drift is as in (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 using the ideas in Delattre et al. (2013). The proofs are a natural extension of their setting to the multidimensional, affine, non-homogeneous case, but become more tedious to work out in detail and to write down and will therefore be omitted here. We will get back to the affine model with i.i.d. observations in subsection 2.3.1 and in section 3.

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 expansion argument together with a required asymptotic normality of the N-sample Score function and a convergence of the average Fisher Information (FI) (see, e.g., 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 (1)) and the standard central limit theorem for i.i.d. variables can not be applied to the Score function, one can revert to the Lindeberg-Feller central limit theorem, given 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 i.i.d. models, often breaks down to requiring that covariate averages converge (Fahrmeir and Kaufmann, 1985).

The more general LAN approach which we pursue here dispenses with the differentiability conditions by building upon L_2-derivatives. An L_2-Score function and L_2-Fisher information are defined, which then are required to meet the above mentioned Lindeberg and convergence conditions (cf. assumption (e) below and Theorem 3). The first part of this section adapts results 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 a second part, we illustrate the verification of regularity conditions for an SDMEM with covariates and with dynamics that are frequently

encountered in biomedical modeling. While the L_2-based approach opens up for the inclusion of irregular densities into our framework, it still requires one to verify the convergence of the average FI. We will discuss the complications of the latter within the SDMEM framework at the end of this section.

We write $\nu^i = \mathbb{Q}^i_{\mu_0,\varphi_0}$ (see 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-a.s. 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).
(c) $\psi^i(\theta)$ is continuous in $L_2(\nu^i)$.
As a consequence, the matrix $I^i(\theta) = 4 \int \psi^i(\theta; x)' \psi^i(\theta; x) d\nu^i(x)$ exists and is continuous and will be called the FI matrix. The N-sample FI is then $I_N(\theta) = \sum_{i=1}^N I^i(\theta)$.
(d) The FI is bounded away from 0 and finite: $0 < \inf_{\theta \in \Theta} [\![\frac{1}{N} I_N(\theta)]\!] \leq \sup_{\theta \in \Theta} [\![\frac{1}{N} I_N(\theta)]\!] < \infty$.
(e) There is a symmetric, positive definite limiting matrix $I(\theta)$ such that
$$\lim_{N\to\infty} \sup_{\theta \in K} [\![\tfrac{1}{N} I_N(\theta) - I(\theta)]\!] = 0 \text{ and } \lim_{N\to\infty} \sup_{\theta \in K} [\![(\tfrac{1}{N} I_N(\theta))^{-1/2} - I(\theta)^{-1/2}]\!] = 0.$$

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 $S_N(\theta) = \sum_{i=1}^N S^i(\theta)$ for the N-sample score function. One can show that also in this more general setting the score function is centered (Ibragimov and Has'minskii, 2013, p. 115).

Sufficient conditions for the a.s. 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 F is continuous and assume for simplicity $\Sigma(t, x) \equiv I$ is the identity matrix. If $\mu \mapsto F(X_s^i, D_s^i, \mu, \varphi)$ is uniformly continuous (for instance differentiable with bounded Jacobian), then $\mu \mapsto \int_0^{T^i} F(X_s^i, D_s^i, \mu, \varphi)' F(X_s^i, D_s^i, \mu, \varphi) ds$ is continuous. If F moreover has the property $\|F(X^i, D_s^i, \mu, \varphi) - F(X^i, D_s^i, \mu_0, \varphi)\| \leq K(1 + \|X^i\|^{\kappa})\|\mu - \mu_0\|$ for some $\kappa > 0$, Kolmogorov's continuity criterion guarantees continuity of q^i.

The L_2-differentiability is neither stronger nor weaker than standard (point-wise) 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 $\psi^i(\theta; x) = \frac{d}{d\theta}[p^i(\theta; x)^{1/2}]$.

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 $\tilde{S}^i(\theta) := 2p^i(\theta)^{-1/2}\frac{d}{d\theta}p^i(\theta)$ 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 point-wise derivative and in fact, $\tilde{S}^i(\theta) = S^i(\theta)$ and $\tilde{I}^i(\theta) = I^i(\theta)$.

---

†For each θ, $\int \|\psi^i(\theta; x)\|^2 d\nu^i(x) < \infty$ and $\lim_{\|h\|\to 0}\|h\|^{-2} \int \|\sqrt{p^i(\theta + h; x)} - \sqrt{p^i(\theta; x)} - \psi^i(\theta; x)h\|^2 d\nu^i(x) = 0$.

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

In the sequel, we write shortly 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 are now in the position to state results on the asymptotic behavior of the MLE in SDMEMs with covariates. These are consequences of theorems in Ibragimov and Has'minskii (2013), and proofs are only shortly outlined.

THEOREM 2 (CONSISTENCY). *The MLE of model (1) is uniformly on K consistent, if*

(A.1) There is a constant $m > q$ such that $\sup_{\theta \in K} \mathbb{E}_\theta \left(\|S_N(\theta)\|^m \right) < \infty$.

(A.2) There is a positive constant $a(K)$ such that for (sufficiently large N and) all $\theta \in K$ (and all
$h \in \Theta_{N,\theta}$) $H_i^2(\theta, \theta_N) \geq a(K)\frac{\|\theta_N - \theta\|^2}{1 + \|\theta_N - \theta\|^2}$, *where $H_i^2(\theta_1, \theta_2) := \int \left(\sqrt{p^i(\theta_1)} - \sqrt{p^i(\theta_2)} \right)^2 d\nu^i$*
is the squared Hellinger distance between $\mathbb{Q}_{\theta_1}^i$ and $\mathbb{Q}_{\theta_2}^i$.

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

REMARK 2. *If the dimension of the parameter set is 1, (A.1) can be replaced by a sub-quadratic growth condition on the Hellinger distance (for i.i.d. observations, see Ibragimov and Has'minskii (2013, Theorem I.5.3)), namely that $H^2(\theta_1, \theta_2) \leq A\|\theta_2 - \theta_1\|^2$, 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 (Thms II.6.2. and III.1.1, Ibragimov and Has'minskii (2013)).

THEOREM 3 (ASYMPTOTIC NORMALITY). *Assume (A.1) and (A.2) from Theorem 2 and additionally*

(B.1) The family $\{S^i(\theta), i = 1, \dots, N\}$ satisfies the Lyapunov condition uniformly in K, i.e.,
there is $\delta > 0$ such that $\lim_{N \to \infty} \sup_{\theta \in K} \sum_{i=1}^N \mathbb{E}_\theta \left(\left\| I_N(\theta)^{-1/2}S^i(\theta) \right\|^{2+\delta} \right) = 0$.

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

Then $\{\hat{\theta}_N\}_{N \in \mathbb{N}}$ is uniformly in K consistent, 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 (B.1) can be generalized to the Lindeberg condition. If the densities $\sqrt{p^i(\theta)}$ are twice continuously differentiable with second derivative $J^i(\theta)$, (B.2) can be replaced by requiring that $\lim_{N \to \infty} \sup_{\theta \in K} \sup_{\|h\| \leq R} [\![I_N(\theta)^{-1/2}]\!]^4 \sum_{i=1}^N \int [\![J^i(\theta_N)]\!]^2 d\nu^i = 0$. As pointed out in the introduction, 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 $\log p^i(\theta) = \log \left(\int q^i(\mu, \varphi)g(\varphi; \vartheta)d\varphi \right)$. 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, e.g., include differentiability of q^i w.r.t. μ and a uniform in μ domination of $\frac{d}{d\mu}q^i(\mu, \varphi)(\int q^i(\mu, \varphi)g(\varphi; \vartheta)d\varphi)^{-1}$. 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.

2.3.1.   SDMEM with covariates and affine mixed effects

We illustrate the verification of the assumptions for Theorems 2 and 3 for the affine SDMEM (2). This model, which certainly is more regular than required, will be revisited in section 3, where we study estimation performance and hypothesis testing for different sample sizes and sampling frequencies, and will be revisited in section 4 for the statistical investigation of EEG data. For simplicity we assume $B = C$, 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 a.s. invertible), the score function for subject i is thus given by $S^i(\theta) = \left[\frac{d}{d\mu} \log p^i(\theta), \frac{d}{d\Omega} \log p^i(\Omega)'\right]$, with

$$\frac{d}{d\mu} \log p^i(\theta) = \gamma^i(\theta)' \qquad \text{and} \qquad \frac{d}{d\Omega} \log p^i(\theta) = \frac{1}{2}\left[-G^i(\Omega) + \gamma^i(\theta)\gamma^i(\theta)'\right].$$

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

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

$$H^2(\theta, \theta_N) = \int \left[-\psi_N(\theta)(\theta_N - \theta) + \left(\sqrt{p_N(\theta_N)} - \sqrt{p_N(\theta)}\right) + \psi_N(\theta)(\theta_N - \theta)\right]^2 d\nu$$

$$= \int \left[-\psi_N(\theta)(\theta_N - \theta)\right]^2 d\nu + o(\|\theta_N - \theta\|^2)$$

$$= (\theta_N - \theta)' I_N(\theta)(\theta_N - \theta) + o(\|\theta_N - \theta\|^2) - 2O(\|\theta_N - \theta\|^2)o(\|\theta_N - \theta\|^2)$$

$$\geq \|(\theta_N - \theta)\|^2 \lambda_{N,\min}(\theta) + o(\|\theta_N - \theta\|^2).$$

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 $H^2(\theta, \theta_N) \geq A_K \|(\theta_N - \theta)\|^2$. Since Θ is bounded, we even have $\|(\theta_N - \theta)\|^2 \geq C \frac{\|(\theta_N - \theta)\|^2}{1 + \|(\theta_N - \theta)\|^2}$ for some positive constant C, which shows that (A.2) holds.

The Lyapunov condition (B.1) follows in a straightforward way. According to the above, $\mathbb{E}_\theta(\|S^i(\theta)\|^3) \leq C$ for some C and therefore

$$\sup_{\theta \in K} \sum_{i=1}^N \mathbb{E}_\theta\left(\|I_N(\theta)^{-1/2} S^i(\theta)\|^3\right) \leq N^{-3/2} \sup_{\theta \in K}[\![\sqrt{N} I_N(\theta)^{-1/2} - I(\theta)^{-1/2}]\!] \sum_{i=1}^N \mathbb{E}_\theta\left(\|S^i(\theta)\|^3\right)$$

$$+ N^{-3/2} \sup_{\theta \in K}[\![I(\theta)^{-1/2}]\!] \sum_{i=1}^N \mathbb{E}_\theta\left(\|S^i(\theta)\|^3\right)$$

$$\leq C N^{-1/2}\left[\sup_{\theta \in K}[\![\sqrt{N} I_N(\theta)^{-1/2} - I(\theta)^{-1/2}]\!] + \sup_{\theta \in K}[\![I(\theta)^{-1/2}]\!]\right],$$

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

To verify (B.2), we show that (recall that $J^i(\theta)$ denotes the second derivative of $\sqrt{p^i(\theta)}$)

$$\sup_{\|h\| \leq R} \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\nu^i} \left(\left[\!\left[J^i(\theta_N) - J^i(\theta) \right]\!\right]^2 \right) \right] \quad \text{and} \quad \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\nu^i} \left(\left[\!\left[J^i(\theta) \right]\!\right]^2 \right) \right] \qquad (6)$$

converge to 0 uniformly in K. As $J^i(\theta)$ is continuous, it is uniformly continuous on compacta, such that for all $i \in \mathbb{N}$, $a_{i,N} = \sup_{\|h\| \leq R} \left[\!\left[J^i(\theta_N) - J^i(\theta) \right]\!\right]$ converges a.s. to 0 as $N \to \infty$. One can show that $a_{i,N} \leq A^i(\theta, R)$ and $\mathbb{E}_{\nu^i}\left(A^i(\theta, R)^2 \right) \leq D_K$. Dominated convergence implies $\mathbb{E}_\theta(a_{i,N}) \to 0$, and the uniform (in i) bound D_K implies uniform in K convergence of the left term in (6) to 0. For the right term in (6) we note that $\mathbb{E}_{\nu^i}\left(\left[\!\left[J^i(\theta) \right]\!\right]^2 \right) \leq \mathbb{E}_\theta\left(\left[\!\left[\frac{d}{d\theta} S^i(\theta) \right]\!\right]^2 \right) + \mathbb{E}_\theta\left(\left[\!\left[S^i(\theta)' S^i(\theta) \right]\!\right]^2 \right) < C_K$, where C_K is a constant that only depends on K and we conclude uniform in $\theta \in K$ convergence to 0, completing the verification of (B.2).

2.3.2.   On the convergence of the average Fisher information in SDMEMs

As seen above, a key condition for establishing the asymptotic normality of the MLEs was the convergence of the scaled N-sample Fisher information $\frac{1}{N} I_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} I^i(\theta)$ 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 behavior of the averaged FI is deduced from a comparable asymptotic behavior of the averaged covariates, such that the verification of conditions can conveniently be accomplished on covariate level. Also in SDMEMs with covariates, it would be desirable to be able to break down the convergence of $\frac{1}{N} I_N(\theta)$ to an average covariate behavior. This, however, is not possible, not even if we assume the simplest case where the drift function F is linear in state, covariates, fixed and random effects and if the latter 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 $dX_t^i = \left[X_t^i(\mu^1 + \phi^{i,1}) + D_t^i(\mu^2 + \phi^{i,2}) \right] dt + dW_t^i$, with fixed effects vector $\mu = (\mu^1, \mu^2)'$, i.i.d. $\mathcal{N}(0, \Omega)$-distributed random effects $\phi^i = (\phi^{i,1}, \phi^{i,2})'$, and known covariate process D^i. We assume Ω is known, such that $\theta = \mu$ is the only unknown parameter. This setup is a special case of (2) with $A = 0$, $B = C$ and therefore $U_i := U_{1i} = U_{2i}, V_i := V_{1i} = V_{2i} = Z_i$. More specifically,

$$U_i = \begin{pmatrix} \int_0^T X_t^i dX_t^i \\ \int_0^T D_t^i dX_t^i \end{pmatrix} \quad \text{and} \quad V_i = \begin{pmatrix} \int_0^{T^i} (X_t^i)^2 dt & \int_0^{T^i} X_t^i D_t^i dt \\ \int_0^{T^i} X_t^i D_t^i dt & \int_0^{T^i} (D_t^i)^2 dt \end{pmatrix}.$$

The FI is by definition $I^i(\mu) = \mathbb{E}_\mu \left(-\frac{d^2}{d\mu^2} \log p^i(\theta) \right) = \mathbb{E}_\mu \left((I + V_i \Omega)^{-1} V_i \right)$, see eq. (4). The matrix $(I + V_i \Omega)^{-1} V_i$ 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, 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 $\mathbb{E}_\mu \left(x^i(I + z_i'\Omega z_i)^{-1}(x^i)' \right) = x^i(I + z_i'\Omega z_i)^{-1}(x^i)'$. The crucial difference, as compared to the linear SDMEM case, is that the matrix $(I + z_i'\Omega z_i)^{-1}$ is deterministic. Therefore, convergence of $\frac{1}{N} \sum_{i=1}^{N} I^i(\theta)$ is implied by a limiting behavior of averages. This is particularly attractive as one can often design the experiment in such a way that the required limiting behavior

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 $\frac{1}{N}I_N(\theta) \to I(\theta)$ holds, due to the combination of non-linearity and stochasticity.

2.4. 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 sub-parameter of the fixed effect μ, $1 \leq \ell \leq p$, 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, $1 \leq k \leq \ell$, linear null hypothesis $H_0 : L\beta = \eta_0$, where L is a $k \times \ell$ matrix of rank k, specifying the linear hypotheses of interest, and $\eta_0 \in \mathbb{R}^k$. The Wald test statistic is $\left(L\hat{\beta}_N - \eta_0\right)' \left(L\hat{V}_N L'\right)^{-1} \left(L\hat{\beta}_N - \eta_0\right)$, where $\hat{\beta}_N$ is the MLE of β and $\hat{V}_N = \widehat{\mathbb{C}\text{ov}}(\hat{\beta}_N)$ denotes its 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 (LR) 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 is inspired from a study on the selenomethionine metabolism in humans (Große Ruse et al., 2015). This multidimensional 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 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. Additionally, dynamics may depend on covariates D^i. Here, the $D^i \in \{0, 1\}$ encodes the affinity of subject i to one of two treatment groups. For simplicity we assume a unit diffusion matrix, such that we consider the model

$$dX_t^i = F(X_t^i, D^i, \mu, \phi^i)dt + dW_t^i = -G(\alpha + \phi^i)X_t^i + D^i\beta dt + dW_t^i,$$

for $0 \leq t \leq T$ 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). The (unknown) fixed effect μ has the 6-dimensional component α, which is shared across both groups (*placebo* and *treatment*) and an additional 5-dimensional component β, which describes the effect of the covariate (treatment) on a subject's dynamics. We let $\beta' = (1, 2, 3, 1, -2)$. The random effects ϕ^i are i.i.d. $\mathcal{N}(0, \Omega)$-distributed with unknown Ω, which we assume to be diagonal with entries $\text{diag}(\Omega) = \left(0.5^2, 1^2, 1^2, 0.5^2, 0.3^2, 0.3^2\right)$.

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, while those for individuals in the treatment group have average long-term mean $(G(\alpha))^{-1}\beta = (7.50, 4.25, 5.00, 8.00, 14.00)'$, see also Fig. 2. The observation horizon is fixed to $T = 15$. A trajectory of $(X_t^1, \ldots, X_t^N)_{0 \leq t \leq T}$ is simulated with the Euler-Maruyama scheme with simulation step size $\delta = 10^{-4}$. Fig. 2 shows four realized trajectories of the 5-dim. process X^i. The upper two panels 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 b-th 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 $\Delta t = \delta \cdot b = 0.001, 0.01$. To investigate estimation performance as a function of sample size, we used $N = 20$ and $N = 50$. Estimation for the considered $(\Delta t, N)$-combinations was repeated on $M = 500$ simulated data sets. Table 1 reports the sample estimates of relative biases and root mean squared errors (RMSE) of the fixed effects and of the variances of the random effects. The relative bias of $\hat{\alpha}_j$ is computed as $\frac{1}{M}\sum_{m=1}^{M} \frac{\hat{\alpha}_j^{(m)} - \alpha_j}{\alpha_j}$ and the RMSE as $\left(\frac{1}{M}\sum_{m=1}^{M}(\hat{\alpha}_j^{(m)} - \alpha_j)^2\right)^{1/2}, j = 1, \ldots, 6$, with an analogous definition for the other parameters. The first six rows in the table correspond to estimated biases and RMSEs of the shared fixed effects $\alpha_j, j = 1, \ldots, 6$. 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 is 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, 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 $\Delta t = 0.1$. Simulations have shown 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%, as compared to the results 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 subsection 2.4) for this model and investigate the test's power under different non-zero treatment effects. The estimated variance-covariance matrix $\hat{V}_N = \widehat{\mathbb{C}\mathrm{ov}}(\hat{\beta}_N)$ 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 (not shown here) reveal that the asymptotic distribution of the

Table 1. Linear transfer model. Shown are estimated relative bias and RMSE of $\hat{\alpha}, \hat{\beta}$ and diag($\hat{\Omega}$). The sample sizes are $N = 20, 50$, and 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.

$(N, \Delta t)$		$(20, 0.001)$		$(50, 0.001)$		$(50, 0.01)$	
true value		rel. bias	RMSE	rel. bias	RMSE	rel. bias	RMSE
	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
α	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
	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
β	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
	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
diag(Ω)	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

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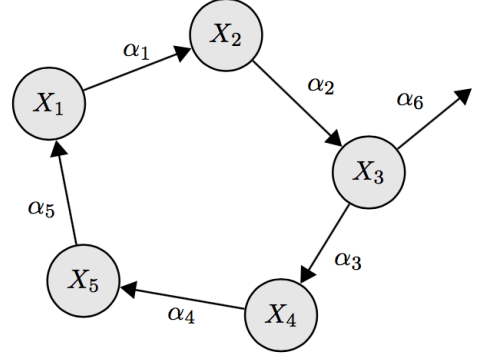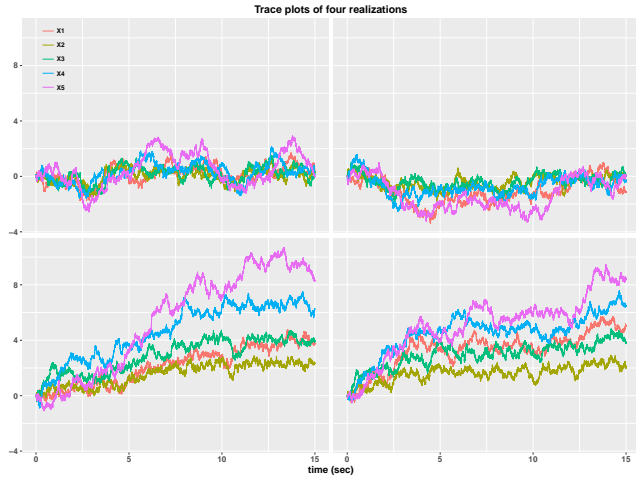

Fig. 2. Linear transfer model: Four realizations of the 5-dimensional state process. The upper two panels show realizations when the covariate is 0 (reference group) and the lower two panels display trajectories for $D^i = 1$ (treatment group). Note the clearly visible difference in the long-term means between the two groups.

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 (based on M under H_0 generated data sets) is 0.074, revealing a slightly liberal finite-sample test behavior. The power of detecting a treatment effect (rejecting $H_0 : \beta = 0$) was computed for different values of β. For $\beta = (1, 2, 3, 1, -2)'$ (values as above), the estimated power was 1. This comes to 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 to 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' 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, e.g., only one coordinate, $\beta = (0.1, 0, 0, 0, 0)'$. 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, while being only slightly conservative, the asymptotic Wald test is able to detect a treatment effect for a rather small data set, even if it causes only a little change of the long-term mean as compared to the standard deviation of the process.

4. Analysis of EEG data

Scalp electroencephalography (EEG) 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 determining a suitable treatment for the patient. The data set was collected during a study conducted by the Children's Hospital Boston and is described in Shoeb (2009). It consists of continuous EEG recordings on 23 epilepsy patients. The electrodes were arranged on the scalp according to the international 10-20 system (see Fig. 3) and the EEG signal was recorded with a sampling frequency of 256 Hz. This is high frequency compared to the typical time scales of the system, and thus, for this type of data, the discretization error will be negligible. During time of recording, every patient experienced one or more periods of abnormal activity that have been classified as epileptic seizures by Shoeb (2009).

Fig. 3. Location of scalp electrodes according to the international 10-20 system

Part of this data set was also analyzed in Østergaard et al. (2017). Their results, which were obtained using a different modeling 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 one 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 located on the left hemisphere and the latter two are, mirrored, on the right. For every patient we extracted two 5s periods of recording, one of them reflecting normal brain activity and the other reflecting abnormal activity classified as epileptic seizure. Fig. 4 shows data for the selected periods of an 11-year old boy. The rows in the plot correspond to the four selected channels and the color indicates pre-seizure (red) and seizure (blue) sections.

The dynamics of the signals during seizure differ clearly from pre-seizure behavior and the objective of this analysis is to better understand, 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 behavior of spontaneous brain activity can be considered stationary. We model the 5s sections of EEG recordings from the four selected channels by a stationary four-dimensional Ornstein-Uhlenbeck (OU) process. This is a process with dynamics $dX_t = AX_t dt + \Sigma dW_t$ and explicit solution $X_t = e^{At}x_0 + \int_0^t e^{A(t-s)}\Sigma dW_s$. In particular, X_t (given x_0) is Gaussian with mean $\mathbb{E}(X_t) = e^{At}x_0$ and covariance matrix $\mathbb{V}(X_t) = \int_0^t e^{As}\Sigma\Sigma'e^{A's}ds$. If all eigenvalues of the rate matrix A have negative real parts, X has a stationary solution and the stationary distribution is a centered Gaussian with covariance matrix $V = \int_0^\infty e^{Au}\Sigma\Sigma'e^{A'u}du$ and autocorrelation function (ACF) $r_X(\tau) = V^{1/2}e^{A'\tau}V^{-1/2}$.
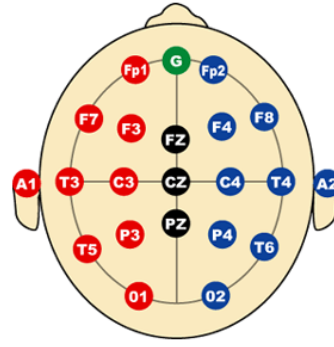
Fig. 4. Sections of $5s$ of the EEG-recording during pre-seizure activity and during an epileptic seizure for the four channels FP1_F7, FP1_F3, FP2_F4 and FP2_F8 and a single subject, an 11-year old boy. The pre-seizure time series is plotted in red and the signal during a subsequent seizure is given in blue.

The statistical model

The prevalent inter-subject variability for EEG 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 EEG data first and afterwards give a motivation for our choice. We denote the pre-seizure process of subject i by $Y^{i,1}$ and the seizure process by $Y^{i,2}$. During seizure, the signal is amplified considerably (Fig. 4). As structural differences are easier to analyze when pre-seizure and seizure data are of comparable magnitude, we re-scale 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 (square root of the quadratic variation) of channel j. Normalizing by a diagonal matrix does not introduce changes to 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 base model, we then model the (normalized) data for subject i by

$$dX_t^{i,k} = \left[A + \Phi^{i,\mathrm{pre}} + D^{i,k}(M + \Phi^{i,\delta}) \right] X_t^{i,k} dt + \Sigma dW_t^{i,k}, \tag{7}$$

where $W^{i,k}$ are independent Brownian motions, $A, M, \Phi^{i,\mathrm{pre}}, \Phi^{i,\delta}$ are 4×4 matrices and the entries of $\Phi^{i,\mathrm{pre}}, \Phi^{i,\delta}$ are independent centered Gaussian random variables (the random effects). The covariate $D^{i,k}$ encodes whether the data belongs to pre-seizure ($D^{i,1} = 0$) or seizure state ($D^{i,2} = 1$). Thus, for a pre-seizure state, population dynamics are driven by the rate matrix A,

whereas M represents the covariate (or seizure) effect. Rewriting equation (7) as

$$dX_t^{i,k} = \left[B(X_t^{i,k}, D^{i,k})\mu + C(X_t^{i,k}, D^{i,k}) \begin{pmatrix} \phi^{i,\mathrm{pre}} \\ \phi^{i,\delta} \end{pmatrix} \right] dt + \Sigma dW_t^{i,k}, \tag{8}$$

(with $\phi^{i,\mathrm{pre}}$ and $\phi^{i,\delta}$ being the vectorized versions of $\Phi^{i,\mathrm{pre}}$ and $\Phi^{i,\delta}$, 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.

Motivation for the model approach

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

Fig. 5. 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. The gray lines correspond to individual estimates for pre-seizure (red) and seizure (blue) data. The black line is the mean correlation, averaged over individuals and states.
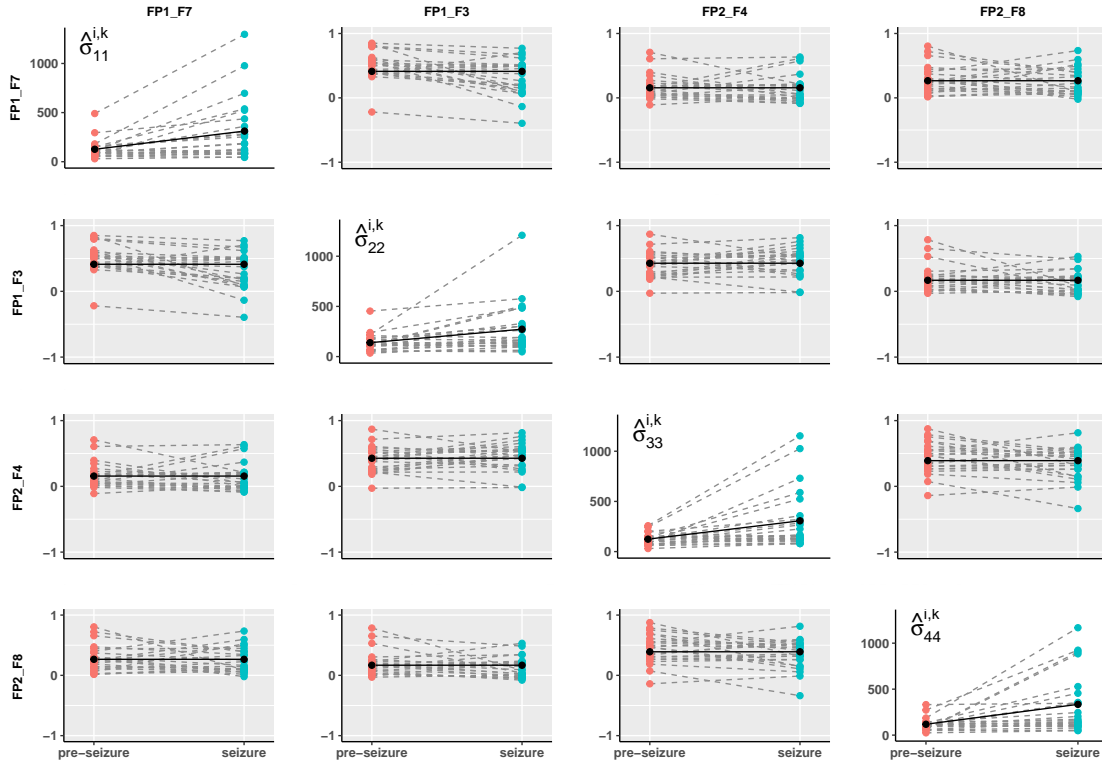

Fig. 5 shows that the average structure of the infinitesimal correlations between channels (off-diagonal plots) does not differ considerably between pre-seizure (red) and seizure states

(blue). 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 pre-seizure 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 pre-seizure to seizure state is modeled in terms of the drift matrix $M + \Phi^{i,\delta}$. 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 $\Phi^{i,\delta}$.

Results

The statistical conclusions are based on the population rate matrices A, M, which are estimated by their MLEs as outlined in section 2.2. The estimates of the population-based rate matrices are

$$
\hat{A} = \begin{pmatrix} -10.52 & -3.59 & -0.42 & 2.47 \\ 3.24 & -17.72 & 4.76 & 1.70 \\ 1.98 & 0.14 & -12.60 & 3.94 \\ 0.74 & -1.75 & -1.52 & -12.87 \end{pmatrix} ; \qquad \hat{M} = \begin{pmatrix} -3.22 & 2.65 & 0.80 & -0.16 \\ 0.83 & 4.60 & -1.51 & 2.81 \\ -0.82 & -0.27 & 0.74 & 0.00 \\ 3.27 & 0.59 & 1.30 & -3.36 \end{pmatrix} .
$$

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

Fig. 6. Confidence intervals for the estimated entries of the covariate effect matrix M. The blue lines are for the full model, whereas the 16 black lines are derived from 16 reduced models where all but one entry of M are set to 0.
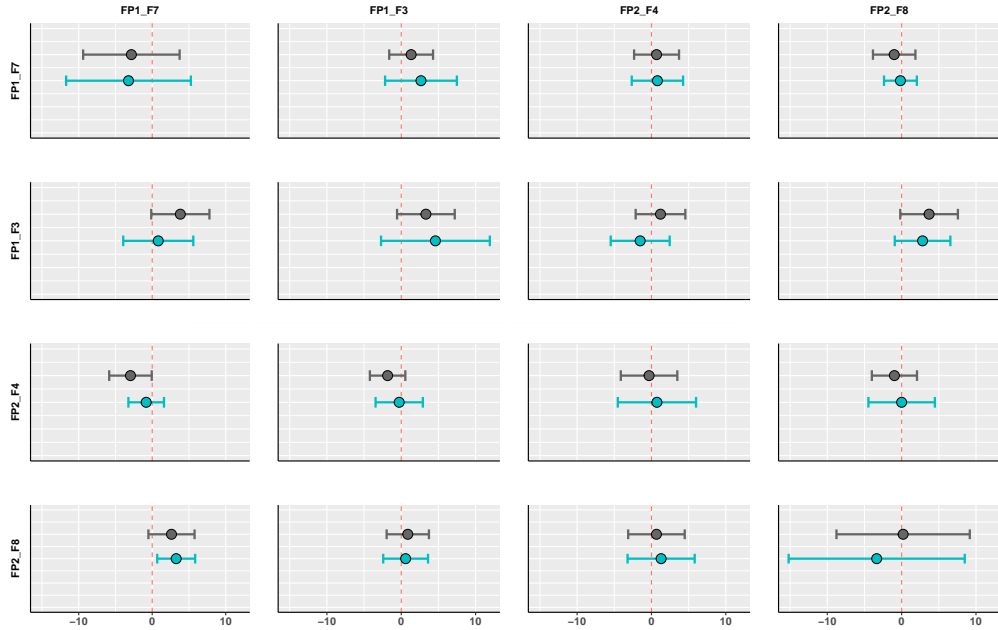
Table 2. Stationary correlations between channels, for pre-seizure state and seizure state (columns 4,5). The last column shows the change in correlation for seizure epochs as compared to non-seizure periods.

| channel 1	channel 2	correlation		change (%)
		pre-seizure	*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

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 hard to detect and statistical results have to be interpreted with caution. More insight into where changes might be present in the rate matrix between pre-seizure and seizure states is provided in Fig. 6. 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. 6 is derived from a reduced model in which all but one elements of M are set to 0. As expected, most CIs are more narrow, however, only few elements appear to have an effect. The lower left plot, e.g., suggests an increased influence of channel FP2_F8 on FP1_F7 under seizure.

It is not straightforward to interpret a covariate effect by looking at the matrix M in an entry-by-entry manner. 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 analyze this, we compare the (population) stationary covariance matrices of pre-seizure 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 pre-seizure 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 ACFs shown in Fig. 7. The diagonal panels in Fig. 7 show the univariate autocorrelation for every channel, i.e., the correlations between a channel and its time-lagged version, as a function of the time lag. The autocorrelations show no marked difference between pre-seizure 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 inverse indicate the typical time constants in the system. For the pre-seizure 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.

Fig. 7. Theoretical multivariate ACFs $r_X^k(\tau) = \hat{V}_k^{1/2}e^{\hat{A}_k'\tau}\hat{V}_k^{-1/2}$, $k = 1, 2$, for the stationary distributions, where $\hat{A}_1 = \hat{A}$ is the estimated population rate matrix for pre-seizure states (in blue), $\hat{A}_2 = \hat{A} + \hat{M}$ is for seizure states (red), and $\hat{V}_k$ is the estimated stationary (population) covariance matrix.
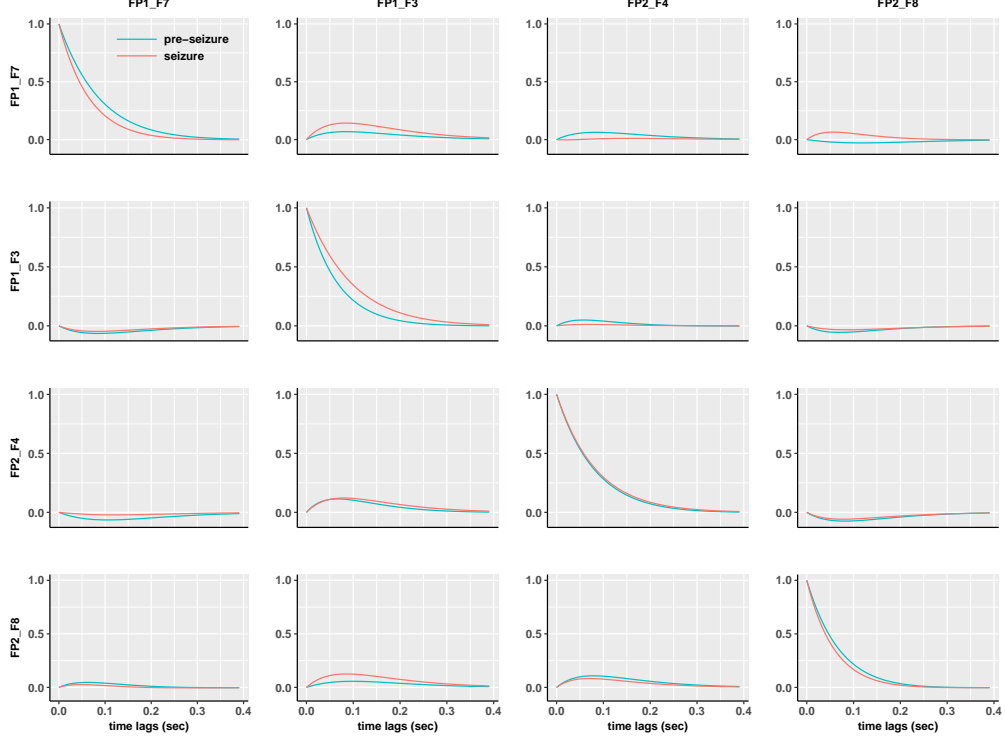
Nonetheless, the main effect of an epileptic seizure seems to be captured by increased variance addressed in the rescaling of the data, more than structural changes. However, with only 23 patients finer effects might be difficult to unravel. An analysis of all channels would be of interest, but is only possible with much larger sample sizes due to the many parameters a full analysis would imply.

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 nonlinear states. The inclusion of (deterministic) subject-specific covariate information, which causes the modeler to leave the world of identically distributed observations, is addressed as well. The presented conditions 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 multidimensional extension of the one studied in Delattre et al. (2013) and is, in its multidimensional 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 introduced by Lorenz (1963), which have been used to model, e.g., temperature, wind speed and humidity, the Brusselator model (Kondepudi and Prigogine, 2014, 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 SIR (susceptible-infected-removed) model 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 convincing already for small sample sizes ($N = 20$), despite the comparably large number (11 fixed effects and 6 variances) of unknown parameters. A moderate sampling interval (of $\Delta t = 0.01$) still gave good results for the considered sample size. 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 LR test. Based on 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 EEG data to assess differences between dynamics for non-seizure and seizure periods. The population voltage dynamics during non-seizure and seizure states are modeled as Ornstein-Uhlenbeck processes, while the prevalent inter-subject variability was accounted for by the inclusion of random effects in the drift. After having adjusted for the subject-specific deviations, systematic differences between pre-seizure 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 as compared to non-seizure states.

A few comments are in order concerning the simulation study and the presented application. Regarding the EEG data analysis, it should be noted that a physiological interpretation of our results in terms of an underlying network structure has to be taken with a grain of salt for two key reasons. One is that EEG 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 only studied the method's applicability to models with up to 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 presented approach 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 EEG recordings, but for low-frequency observations, which sometimes occur in pharmacological applications, a severe bias is introduced, which is to bear in mind in applications. A possible solution is to impute data at time points in between observation times, and 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 coordinates in the state space are observed, and an entire path of a completely unobserved (latent) coordinate should be inferred (Berg and Ditlevsen, 2013; Ditlevsen and Samson, 2014). Missing observations of one or more coordinates is not untypical for biological data. This, at a first step, prohibits application of the proposed estimation procedure, as it relies on the assumption of complete data observations. Such statistical recovery of hidden state coordinates remains a topic for future research.

Acknowledgements

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica 70*(1), 223–262.

Beal, S. L. and L. B. Sheiner (1981). Estimating population kinetics. *Critical Reviews in Biomedical Engineering 8*(3), 195–222.

Berg, R. W. and S. Ditlevsen (2013). Synaptic inhibition and excitation estimated via the time constant of membrane potential fluctuations. *Journal of Neurophysiology 110*(4), 1021–1034.

Bladt, M., S. Finch, and M. Sørensen (2016). Simulation of multivariate diffusion bridges. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(2), 343–369.

Bradley, R. A. and J. J. Gart (1962). The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika 49*(1/2), 205–214.

Delattre, M., V. Genon-Catalot, and C. Laredo (2017). Parametric inference for discrete observations of diffusion processes with mixed effects. *Stochastic Processes and their Applications*.

Delattre, M., V. Genon-Catalot, and A. Samson (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics 40*(2), 322–343.

Delattre, M., V. Genon-Catalot, and A. Samson (2015). Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. *ESAIM: Probability and Statistics 19*, 671–688.

Delyon, B., V. Lavielle, and E. Moulines (1999). Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics 27*(1), 94–128.

Ditlevsen, S. and A. Samson (2014). Estimation in the partially observed stochastic Morris–Lecar neuronal model with particle filter and stochastic approximation methods. *The Annals of Applied Statistics 8*(2), 674–702.

Donnet, S., J.-L. Foulley, and A. Samson (2010). Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics 66*(3), 733–741.

Durham, G. B. and A. R. Gallant (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics 20*(3), 297–338.

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

FitzHugh, R. (1955). Mathematical models of threshold phenomena in the nerve membrane. *The Bulletin of Mathematical Biophysics 17*(4), 257–278.

Große Ruse, M., L. R. Søndergaard, S. Ditlevsen, M. Damgaard, S. Fuglsang, J. T. Ottesen, and J. L. Madsen (2015). Absorption and initial metabolism of 75 se-l-selenomethionine: a kinetic model based on dynamic scintigraphic data. *British Journal of Nutrition 114*(10), 1718–1723.

Guy, R., C. Larédo, and E. Vergu (2015). Approximation of epidemic models by diffusion processes and their statistical inference. *Journal of Mathematical Biology 70*(3), 621–646.

Hoadley, B. (1971). Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of Mathematical Statistics 42*, 1977–1991.

Ibragimov, I. A. and R. Z. Has'minskii (2013). *Statistical Estimation: Asymptotic Theory*, Volume 16. Springer Science.

Jensen, A. C., S. Ditlevsen, M. Kessler, and O. Papaspiliopoulos (2012). Markov chain Monte Carlo approach to parameter estimation in the Fitzhugh-Nagumo model. *Physical Review E 86*(4), 041114.

Keeling, M. J. and P. Rohani (2008). *Modeling infectious diseases in humans and animals.* Princeton University Press.

Kermack, W. O. and A. G. McKendrick (1927). A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Volume 115, pp. 700–721. The Royal Society.

Kessler, M., A. Lindner, and M. Sørensen (2012). *Statistical methods for stochastic differential equations.* CRC Press.

Kondepudi, D. and I. Prigogine (2014). *Modern thermodynamics: from heat engines to dissipative structures.* John Wiley &; Sons.

Le Cam, L. (2012). *Asymptotic methods in statistical decision theory.* Springer Science, New York.

Leander, J., T. Lundh, and M. Jirstrand (2014). Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements. *Mathematical Biosciences 251*, 54–62.

Lehmann, E. L. and J. P. Romano (2006). *Testing statistical hypotheses.* Springer Science & Business Media.

Lo, A. W. (1988). Maximum likelihood estimation of generalized Itô processes with discretely sampled data. *Econometric Theory 4*(2), 231–247.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences 20*(2), 130–141.

Møller, J. B., R. V. Overgaard, H. Madsen, T. Hansen, O. Pedersen, and S. H. Ingwersen (2010). Predictive performance for population models using stochastic differential equations applied on data from an oral glucose tolerance test. *Journal of Pharmacokinetics and Pharmacodynamics 37*(1), 85–98.

Murray, J. D. (2002). *Mathematical Biology I: An Introduction*, Volume 17 of *Interdisciplinary Applied Mathematics.* Springer, New York, NY, USA,.

Nagumo, J., S. Arimoto, and S. Yoshizawa (1962). An active pulse transmission line simulating nerve axon. *Proceedings of the IRE 50*(10), 2061–2070.

Østergaard, J., A. Rahbek, and S. Ditlevsen (2017). Oscillating systems with cointegrated phase processes. *Journal of Mathematical Biology 75*(4), 845–883.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics 22*, 55–71.

Phillips, P. C. and J. Yu (2009). Maximum likelihood and gaussian estimation of continuous time models in finance. In *Handbook of Financial Time Series*, pp. 497–530. Springer, New York.

Picchini, U., A. De Gaetano, and S. Ditlevsen (2010). Stochastic differential mixed-effects models. *Scandinavian Journal of Statistics 37*(1), 67–90.

Picchini, U. and S. Ditlevsen (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis 55*(3), 1426–1444.

Picchini, U., S. Ditlevsen, A. De Gaetano, and P. Lansky (2008). Parameters of the Diffusion Leaky Integrate-and-Fire Neuronal Model for a Slowly Fluctuating Signal. *Neural Computation 20*(11), 2696–2714.

Shoeb, A. H. (2009). *Application of machine learning to epileptic seizure onset detection and treatment*. Ph. D. thesis, Massachusetts Institute of Technology.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge University Press.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika 80*(4), 791–795.