Statistics and Its Interface Volume 0 (2021) 1

Generalized Gaussian time series model for increments of EEG data

Nikolai N. Leonenko*, Željka Salinger, Alla Sikorskii, Nenad Šuvak and Michael Boivin

We propose a new strictly stationary time series model with marginal generalized Gaussian distribution and exponentially decaying autocorrelation function for modeling of increments of electroencephalogram (EEG) data collected from Ugandan children during coma from cerebral malaria. The model inherits its appealing properties from the strictly stationary strong mixing Markovian diffusion with invariant generalized Gaussian distribution (GGD). The GGD parametrization used in this paper comprises some famous light-tailed distributions (e.g., Laplace and Gaussian) and some well known and widely applied heavy-tailed distributions (e.g., Student). Two versions of this model fit to the data from each EEG channel. In the first model, marginal distributions is from the light-tailed GGD sub-family, and the distribution parameters were estimated using quasi-likelihood approach. In the second model, marginal distributions is heavy-tailed (Student), and the tail index was estimated using the approach based on the empirical scaling function. The estimated parameters from models across EEG channels were explored as potential predictors of neurocognitive outcomes of these children 6 months after recovering from illness. Several of these parameters were shown to be important predictors even after controlling for neurocognitive scores immediately following cerebral malaria illness and traditional blood and cerebrospinal fluid biomarkers collected during hospitalization.

AMS 2000 subject classifications: Primary 37M10; secondary 62M10, 62G07, 62J20, 62P10.
Keywords and phrases: Time series, Diffusion process, Diffusion discretization, Generalized Gaussian distribution, Heavy-tailed distribution, Tail index.

1. INTRODUCTION

Cerebral malaria is the most severe neurological complication of infection with a parasite *Plasmodium falciparum*. Over 90% of all cerebral malaria cases occur in sub-Saharan Africa, with over half a million children affected each year. For the review of mathematical models for malaria epidemic we refer to [27], while more recent results could be found in [41].

Even with treatment, $10-20\%$ of affected children die, while survivors sustain brain injury, which may affect subsequent neurodevelopment and cognitive functioning [18]. Identification of factors that can predict the extent of neurocogntive impairment and other outcomes following cerebral malaria illness is an important problem [32, 3]. While evidence-based rehabilitative interventions for survivors are available, resources to administer them are limited across sub-Saharan Africa. Therefore directing these interventions to those in most need, as determined by predictors of subsequent impairment, is key to the most efficient use of the available resources.

Onset of coma is one of the diagnostic differences between cerebral and severe malaria, and electroencephalogram (EEG) is used to monitor child's brain during coma by recording data on electrical neural activity of the brain. Signals are captured by multiple electrodes called channels located over the scalp. A number of models linking the chemical processes within corresponding neurons generating the active potentials have been proposed since 1950s. An overview of these models is given in [36]. Statistical analyses of EEG data included classification and prediction using arrays of EEG features, but few models for the underlying stochastic processes have been proposed [33, 40]. In this paper we build upon past work [40] in which it was determined that the underlying stochastic processes across channels are not stationary but have stationary increments. The distributions of these increments were symmetrical across channels; in some channels the distributions were heavy-tailed whereas in other channels, lighter tails were observed. In this paper we provide a unifying stochastic model that captures different types of tails across the range of model parameters. This model is based on time series with marginal Generalized Gaussian Distribution (GGD) and exponentially decaying correlation function.

The paper is structured as follows. In Section 2 we introduce the family of GGD, state its main properties and discuss important special cases. The time series of observations from strictly stationary strong mixing Markovian diffusion with marginal GGD is introduced in Section 3, where we also discuss some alternative models and different schemes for discretization of diffusions. Methods for parameter estimation, quasi-likelihood and a relatively new method for

*Corresponding author.

tail-index estimation, are described in Section 4. Section 5 is dedicated to fitting of the proposed time series models to the increments of real EEG recordings. The first set of models is for non-heavy tailed distributions with parameter estimated by the quasi-likelihood method based on the invariant distribution of the underlying Markovian diffusion. In the second set of models (Section 5.3), marginal heavy-tailed Student distribution of the time series is used, and its tail index is estimated by the method based on the partition function. In Section 6 the elastic net is used to determine whether the parameters of from these two sets of models are useful as predictors of children's neurocognitive scores 6 months after surviving cerebral malaria. The results are discussed in Section 7.

2. DISTRIBUTIONAL PROPERTIES OF EEG INCREMENTS

Our approach in dealing with EEG signal is based on its transformation into EEG increments, which have a symmetric distribution with the maximum at zero. To reflect the diversity of the empirically observed distribution candidates, we parametrize the probability density function (PDF) of the distribution of increments as follows:

$$(1)\quad f_{s,b}(x) = \begin{cases} \dfrac{1}{2(s\sigma^2)^{1/s}\Gamma\left(1+\frac{1}{s}\right)} \exp\left(-\dfrac{|x|^s}{s\sigma^2}\right), & b=0, \\ \dfrac{bs}{2\sigma^2}\left(\dfrac{s\sigma^2}{b}\right)^{-1/s} \dfrac{\Gamma\left(1+\frac{1}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)} \\ \qquad \cdot \left(1+\dfrac{b}{s\sigma^2}|x|^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1}, & b>0, \end{cases}$$

where the value of parameter b is used as an indicator for making a distinction between light-tailed ($b = 0$) and heavy-tiled ($b > 0$) distributions within this family. We will refer to the distributions from this family as GGD. The subfamily characterized by $b = 0$ resembles the usual GGD parametrization including, for $s = 2$, the zero-mean normal distribution with variance σ^2. For more details on light-tailed GGD subfamily we refer to [29] and [11]. For $b > 0$, distributions in the GGD subfamily admit heavy tails, e.g., for $s = 2$ this distribution is of the Student type. For more information on Student distribution and related processes we refer to [14], whereas for parametrization similar to (1) we refer to [25], where this distribution was studied in the framework of affine moments to prove some sharp moment-entropy inequalities. Namely, it is known that the GGD satisfies the maximum entropy principle (see e.g. [26]), which we state here in order to emphasize the importance of this family of distributions to the entropy-based problems.

Theorem 2.1. *Let X be a real random variable with the PDF g such that for some $s > 0$*

$$E\left[|X|^s\right] < \infty,$$

and let $H_q(g)$ be the corresponding Shanon ($q = 1$) or Rènyi ($1/(1+s) < q < 1$) entropy.

(i) If $\operatorname{supp}(g) = \mathbb{R}$ and $q = 1$, then

$$H_1(g) \le H_1(f_{s,0}) = \frac{1}{s} - \log\left(\frac{1}{2(s\sigma^2)^{1/s}\Gamma\left(1+\frac{1}{s}\right)}\right),$$

where $f_{s,0}$ is the PDF of GGD (1) for $b = 0$ and

$$H_1(g) = -\int_{\mathbb{R}} g(x)\log\left(g(x)\right)dx,$$

is the corresponding Shannon entropy. The equality holds if and only if $g = f_{s,0}$ a.s.

(ii) If $\operatorname{supp}(g) = \mathbb{R}$ and $1/(1+s) < q < 1$, then

$$H_q(g) \le H_q(f_{s,b}) = \frac{1}{1-q}\log\left(1+\frac{b}{s}\right) - \log\left(\frac{bs}{2\sigma^2}\left(\frac{s\sigma^2}{b}\right)^{-1/s}\frac{\Gamma\left(1+\frac{1}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)}\right),$$

where $f_{s,b}$ is the PDF of GGD (1) for $b > 0$ and

$$H_q(g) = \frac{1}{1-q}\log\left(\int_{\mathbb{R}} g^q(x)\,dx\right),$$

is the corresponding Rènyi entropy. The equality holds if and only if $g = f_{s,b}$ a.s.

For more details on Rènyi and Shannon entropy and their relations with GGD (1) and its special cases we refer to [25], [16] and [19].

Further, we state some important properties related to moments of the GGD (1). For $b = 0$, all moments of the distribution (1) exist. In this case the absolute moment of order $\nu > 0$ is of the following form:

$$(2)\qquad \mathrm{E}[|X|^\nu] = (s\sigma^2)^{\nu/s}\frac{\Gamma\left(\frac{1+\nu}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}.$$

In particular, integer moments are given by

$$(3)\qquad \mathrm{E}[X^n] = \frac{1+(-1)^n}{2}(s\sigma^2)^{n/s}\frac{\Gamma\left(\frac{1+n}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}, \quad n \in \mathbb{N},$$

and therefore the first four moments are

$$E[X] = E[X^3] = 0,$$
$$E[X^2] = \frac{\left(s\sigma^2\right)^{2/s}\Gamma\left(\frac{3}{s}\right)}{\Gamma\left(\frac{1}{s}\right)},$$
$$E[X^4] = \frac{\left(s\sigma^2\right)^{4/s}\Gamma\left(\frac{5}{s}\right)}{\Gamma\left(\frac{1}{s}\right)}.$$

For heavy-tailed GGD subfamily ($b > 0$), the tail of the density decreases like $|x|^{-1-s\left(\frac{\sigma^2}{b}+1\right)}$. The absolute moment of order $\nu > 0$ exists for $\nu < s\left(\frac{\sigma^2}{b}+1\right)$ and is given by the following expression:

$$\mathrm{E}[|X|^\nu] = \left(\frac{b}{s\sigma^2}\right)^{-\nu/s} \frac{\Gamma\left(\frac{1+\nu}{s}\right)\Gamma\left(1-\frac{\nu}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(1+\frac{\sigma^2}{b}\right)}. \tag{4}$$

The integer moments are given by

$$\mathrm{E}[X^n] = \frac{s\left(1+(-1)^n\right)}{2}\left(\frac{b}{s\sigma^2}\right)^{1-n/s} \cdot \frac{\Gamma\left(\frac{1+n}{s}\right)\Gamma\left(1-\frac{n}{s}+\frac{\sigma^2}{b}\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2}{b}\right)}, \quad n < s\left(\frac{\sigma^2}{b}+1\right). \tag{5}$$

The first six moments are of the following form:

$$E[X] = E[X^3] = E[X^5] = 0,$$

$$E[X^2] = \frac{\Gamma\left(\frac{3}{s}\right)\left(\frac{b}{s\sigma^2}\right)^{-2/s}\Gamma\left(\frac{\sigma^2}{b}-\frac{2}{s}+1\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2+b}{b}\right)},$$

$$E[X^4] = \frac{\Gamma\left(\frac{5}{s}\right)\left(\frac{b}{s\sigma^2}\right)^{-4/s}\Gamma\left(\frac{\sigma^2}{b}-\frac{4}{s}+1\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2+b}{b}\right)},$$

$$E[X^6] = \frac{\Gamma\left(\frac{7}{s}\right)\left(\frac{b}{s\sigma^2}\right)^{-6/s}\Gamma\left(\frac{\sigma^2}{b}-\frac{6}{s}+1\right)}{\Gamma\left(\frac{1}{s}\right)\Gamma\left(\frac{\sigma^2+b}{b}\right)}.$$

Remark 2.1. Beside time series with light-tailed GGD distribution for modeling of the EEG increments, we will also use the special case of the heavy-tailed part of (1), where $b > 0$. As already stated, in that case for $s = 2$ we obtain the Student-type distribution with the PDF

$$f_{2,b}(x) = \frac{\Gamma\left(\frac{\sigma^2}{b}+\frac{3}{2}\right)}{\sqrt{\frac{2\pi\sigma^2}{b}}\Gamma\left(\frac{\sigma^2}{b}+1\right)}\left(1+\frac{b}{2\sigma^2}x^2\right)^{-\frac{\sigma^2}{b}-\frac{3}{2}}, \quad b > 0. \tag{6}$$

The absolute moment of order $\nu > 0$ exists for $\nu < 2\left(\frac{\sigma^2}{b}+1\right)$ and is given by the following expression:

$$\mathrm{E}[|X|^\nu] = \left(\frac{b}{2\sigma^2}\right)^{-\nu/2} \frac{\Gamma\left(\frac{1+\nu}{2}\right)\Gamma\left(1-\frac{\nu}{2}+\frac{\sigma^2}{b}\right)}{\sqrt{\pi}\,\Gamma\left(1+\frac{\sigma^2}{b}\right)}. \tag{7}$$

The integer moments are given by

$$\mathrm{E}[X^n] = \left(1+(-1)^n\right)\left(\frac{b}{2\sigma^2}\right)^{1-n/2} \cdot \frac{\Gamma\left(\frac{1+n}{2}\right)\Gamma\left(1-\frac{n}{2}+\frac{\sigma^2}{b}\right)}{\sqrt{2\pi}\Gamma\left(\frac{\sigma^2}{b}\right)}, \quad n < 2\left(\frac{\sigma^2}{b}+1\right). \tag{8}$$

and the first six moments are of the following form:

$$E[X] = E[X^3] = E[X^5] = 0,$$
$$E[X^2] = 1,$$
$$E[X^4] = \frac{3\sigma^2}{\sigma^2-b},$$
$$E[X^6] = \frac{15\sigma^4}{\sigma^4-3b\sigma^2+2b^2}.$$

For example, the explicit expressions for moments enable the straightforward statistical analysis of strictly stationary diffusion with Student marginals. For more details on analysis of probabilistic properties, moment-based parameter estimation and testing of statistical hypothesis regarding such diffusion we refer to [24].

3. TIME SERIES REPRESENTATION OF THE EEG INCREMENTS

In order to characterize important probabilistic properties of increments including their dependence structure, we view the EEG increments as the time series $(X_n,\, n \in \mathbb{N})$, representing the model for discrete-time observations from the diffusion process $(X_t,\, t \geq 0)$ with the stationary PDF (1). Since the PDF (1) is continuous, bounded, and strictly positive on the whole $\mathbb{R}$, according to the [2][Theorem 2.1, page 193] the stochastic differential equation (SDE)

$$dX_t = -\theta X_t\, dt + \sqrt{v(X_t)}\, dB_t, \quad \theta > 0, \quad t \geq 0, \tag{9}$$

driven by the standard Brownian motion $(B_t,\, t \geq 0)$, where the drift reflects the mean reversion of the process to zero and the diffusion coefficient v is obtained as

$$v(x) = \frac{2\theta}{f(x)} \int_{-\infty}^{x} (-y) f(y)\, dy,$$

admits the unique weak ergodic solution and defines the diffusion with stationary distribution (1), which we call the generalized Gaussian diffusion (GGDiff). The important properties of GGDiff are:

- If X_0 has pdf $f_{s,b}$, the GGDiff is a strictly stationary process;
- If $E[X_t^2] < \infty$, the autocorrelation function of the GGDiff is given by $\mathrm{Corr}(X_s, X_t) = e^{-\theta|t-s|}$;
- According to [10][Proposition 3, Page 115], GGDiff is a strong mixing process with exponentially decaying mixing coefficient $\alpha_X(t) = \sup_{s\geq 0} \alpha(\mathcal{F}_s, \mathcal{F}^{s+t})$, where

$$\alpha(\mathcal{F}_s, \mathcal{F}^{s+t}) = \sup_{A\in\mathcal{F}_s, B\in\mathcal{F}^{s+t}} |P(A\cap B) - P(A)P(B)|,$$

$$A \in \mathcal{F}_s = \sigma\{X_u,\, u \leq s\}, \quad B \in \mathcal{F}^{s+t} = \sigma\{X_u,\, u \geq s+t\}.$$

The time series $(X_n,\, n \in \mathbb{N})$ inherits the strong stationarity, autocorrelation structure, and the strong mixing property of the GGDiff (9), making it a reasonable model for EEG increments and enabling statistical analysis of the EEG data.

Remark 3.1. The diffusion (9) has the linear drift governed by the autocorrelation parameter θ. The diffusion coefficient defined as follows. For the light-tailed case ($b = 0$), the diffusion coefficient is given by

$$
\begin{aligned}
(10)\quad v(x) = &\begin{cases} 2\theta \exp\left(\frac{(-x)^s}{s\sigma^2}\right)\left(\frac{\left(s\sigma^2\right)^{2/s}}{s}\Gamma\left(\frac{2}{s}\right)\right. & \\ \qquad\qquad \left. - \int\limits_x^0 y e^{-\frac{y^s}{s\sigma^2}}\, dy\right), & x < 0 \\ 2\theta \exp\left(\frac{x^s}{s\sigma^2}\right)\left(\frac{\left(s\sigma^2\right)^{2/s}}{s}\Gamma\left(\frac{2}{s}\right)\right. & \\ \qquad\qquad \left. + \int\limits_0^x y e^{-\frac{y^s}{s\sigma^2}}\, dy\right), & x \geq 0 \end{cases} \\
&= 2\theta \exp\left(\frac{(x\cdot \operatorname{sgn}(x))^s}{s\sigma^2}\right)\left(\frac{(s\sigma^2)^{2/s}}{s}\Gamma\left(\frac{2}{s}\right)\right. \\
&\left. + \sigma^{\frac{4}{s}} s^{\frac{2}{s}-1}\left(\gamma\left(\frac{2}{s}, \frac{x^s}{s\sigma^2}\right) - \gamma\left(\frac{2}{s}, \frac{1}{s\sigma^2}\right)\right)\right),
\end{aligned}
$$

where

$$
\gamma(a, x) = \int\limits_0^x t^{a-1} e^{-t}\, dt,
$$

is the lower incomplete gamma function. For $b > 0$ the diffusion coefficient takes the following form:

(11)

$$
\begin{aligned}
v(x) = &\begin{cases} 2\theta\left(1 + \frac{b}{s\sigma^2}(-x)^s\right)^{\frac{\sigma^2}{b}+\frac{1}{s}+1}\left(\frac{1}{s}\left(\frac{\sigma^2 s}{b}\right)^{2/s}\right. & \\ \quad \left. \cdot B\left(\frac{2}{s}, 1 - \frac{1}{s} + \frac{\sigma^2}{b}\right)\right. & \\ \quad \left. - \int\limits_x^0 y\left(1 + \frac{b}{s\sigma^2} y^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1} dy\right), & x < 0 \\ 2\theta\left(1 + \frac{b}{s\sigma^2} x^s\right)^{\frac{\sigma^2}{b}+\frac{1}{s}+1}\left(\frac{1}{s}\left(\frac{\sigma^2 s}{b}\right)^{2/s}\right. & \\ \quad \left. \cdot B\left(\frac{2}{s}, 1 - \frac{1}{s} + \frac{\sigma^2}{b}\right)\right. & \\ \quad \left. + \int\limits_0^x y\left(1 + \frac{b}{s\sigma^2} y^s\right)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1} dy\right), & x \geq 0 \end{cases} \\
&= 2\theta\left(1 + \frac{b}{s\sigma^2}\left(x\cdot\operatorname{sgn}(x)\right)^s\right)^{\frac{\sigma^2}{b}+\frac{1}{s}+1} \\
&\quad \cdot \left(\frac{1}{s}\left(\frac{s\sigma^2}{b}\right)^{\frac{2}{s}} B\left(\frac{2}{s}, 1 - \frac{1}{s} + \frac{\sigma^2}{b}\right) + s^{\frac{2}{s}-1}\left(\frac{\sigma^2}{b}\right)^{\frac{2}{s}}\right. \\
&\quad \left.\left(\beta\left(\frac{bx^s}{s\sigma^2}; \frac{1}{s}, -\frac{\sigma^2}{b} - \frac{1}{s}\right) - \beta\left(\frac{b}{s\sigma^2}; \frac{1}{s}, -\frac{\sigma^2}{b} - \frac{1}{s}\right)\right)\right),
\end{aligned}
$$

where $B(\cdot,\cdot)$ is the standard beta function and

$$
\beta\left(x; b, s, \sigma\right) = \int\limits_{\frac{b}{s\sigma^2}}^{\frac{bx^s}{s\sigma^2}} t^{\frac{1}{s}-1}(1+t)^{-\frac{\sigma^2}{b}-\frac{1}{s}-1}\, dt.
$$

The existence of the strong solution of the SDE (9) for $b > 0$ and $s = 2$ follows from the analysis of generally parametrized Student diffusion in [24]. However, due to the nature of the diffusion parameter for $b = 0$ and general $b > 0$, in this case we verified the existence of a unique strong solution of the GGDiff (9) just up to the explosion time $T(X_0)$. In practical application to the EEG data, this explosion time may correspond to the end of coma.

Remark 3.2. There are many schemes for diffusion discretization, see, for example, [17] or [22] for a detailed exposition on Euler and Milstein schemes for numerical solutions to SDEs. However, the main problem with the time series obtained by these discretization schemes, comprising some form of autoregressive structure, is the lack of the strict stationarity. The detailed exposition on this matter is given in [31], where on pages $55 - 56$ it is stated that the stationary time series can be obtained for diffusions with linear drift and unit volatility. For general non-linear diffusions the transformation of the diffusion to the diffusion with the unit volatility and the local linearization of the drift is proposed. Furthermore, it is shown that the discretization of this transformed diffusion under some technical assumptions (Theorems 3.1 and 3.2) yields the non-explosive and ergodic time-series that converges to the unit-volatility diffusion. For example regarding the diffusion with marginal Student distribution we refer to [31][Example 6, pages $69 - 70$].

Remark 3.3. According to [16][Theorem 3.2] there exists the strictly stationary Ornstein-Uhlenbeck type process $(X_t,\, t \in \mathbb{R})$

$$
\begin{aligned}
X_t &= e^{-\lambda t} X_0 + e^{-\lambda t} \int\limits_0^t e^{\lambda s} dY(\lambda s) \\
&= \int\limits_{-\infty}^t e^{-\lambda(t-s)} dY(\lambda s), \quad \lambda > 0,
\end{aligned}
$$

with marginal Student $T(\nu, \delta, \mu)$ distribution with the PDF

$$f(x)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\delta\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\left(\frac{x-\mu}{\delta}\right)^2\right)^{-\frac{\nu+1}{2}},$$
$$\nu>0,\quad \delta>0,\quad \mu\in\mathbb{R},\quad x\in\mathbb{R},$$

governed by the background driving Levy process $(Y_t,\, t\in\mathbb{R})$. Due to the self-decomposability of the Student distribution, there also exists the strictly stationary solution of the autoregressive equation

$$X_n=cX_{n-1}+\varepsilon_n,\quad n\in\mathbb{N}, \tag{12}$$

where $c\in(0,1)$ and $(\varepsilon_n,\, n\in\mathbb{N})$ is a sequence of IID random variables (the innovation process) independent of the process $(X_n,\, n\in\mathbb{N})$. Also, we have

$$X_0\overset{d}{=}cX_0+\varepsilon_1\sim T(\nu,\delta,\mu),$$

and so, if the relation (12) holds for every $c\in(0,1)$, it follows that $T(\nu,\delta,\mu)$ can be observed as a marginal distribution of the autoregressive time series $(X_n,\, n\in\mathbb{N}_0)$. Furthermore, according to [16][Remark 3.1] and [20][Proposition 2],

$$\varepsilon_1\overset{d}{=}\int_0^{-\log(c)}e^{-s}\,dY(s)\overset{d}{=}Y(1),$$

with the distribution having the cumulant transform

$$\begin{aligned}\kappa_{Y(1)}(0)&=0,\\ \kappa_{Y(1)}(z)&=\log E\left[e^{izY(1)}\right]\\ &=iz\mu-\delta|z|\frac{K_{\nu/2-1}(\delta|z|)}{K_{\nu/2}(\delta|z|)},\quad z\in\mathbb{R}\setminus\{0\},\end{aligned}$$

where

$$K_\lambda(x)=\frac{1}{2}\int_0^\infty u^{\lambda-1}\exp\left\{\frac{1}{2}x\left(u+\frac{1}{u}\right)\right\}du,\quad x>0,$$

is the modified Bessel function of the third kind with the index $\lambda\in\mathbb{R}$. For fixed $\lambda>0$ this function is positive, decreasing and, as $x\to 0+$,

$$K_\lambda(x)\sim\Gamma(\lambda)2^{\lambda-1}x^{-\lambda}.$$

Furthermore, the autoregressive time series $(X_n,\, n\in\mathbb{N}_0)$ has the following important properties:

- $X_n\sim T(\nu,\delta,\mu)$, $\forall n\in\mathbb{N}_0$
- $E[X_n]=\mu$ for $\nu>1$
- $\mathrm{Corr}\,(X_n,X_{n+\tau})=c^{|\tau|}$ for $\nu>2$, $c\in(0,1)$ and $\tau=0,\pm1,\pm2,\ldots$.

Since the heavy-tailed subfamily of GGD (1) for $s=2$ comes down to the Student-type distribution, in this case the $AR(1)$ process constructed in such a way could be used as a time-series model for EEG signals. This $AR(1)$ time series is used for modeling the EEG signals in [40]. Furthermore, in light-tailed ($b=0$) case of the GGD (1) for $s\in(0,1]\cup\{2\}$ the strictly-stationary $AR(1)$ time series with GGD (1) marginals could be constructed since for these values of parameter s this distribution is infinitely-divisible and self-decomposable, see [11][Theorems 5 and 6]. However, due to lack of these properties for $s\in(1,2)$ in $b=0$ case, strictly stationary $AR(1)$ time-series with GGD (1) marginals with arbitrary parameter values cannot be constructed. Therefore, model presented in this paper, based on completely different type of underlying stochastic process, is more general.

4. PARAMETER ESTIMATION

4.1 Quasi-likelihood estimation

In the light-tailed case ($b=0$), the two-dimensional parameter $\zeta=(s,\sigma^2)$ of the stationary distribution of the GGDiff $X=(X_t,\, t\geq 0)$ is estimated by the quasi-likelihood method.

For $\Delta>0$ and Markovian diffusion X, let

$$p_X\left(\Delta,x|x_0;\zeta\right)=\frac{d}{dx}P\left(X_{t+\Delta}\in dx|X_t=x_0\right)$$

be the conditional PDF of $X_{t+\Delta}$ given $\{X_t=x_0\}$. Due to the Markovian structure of GGDiff X, the corresponding likelihood function based on the time series of observations $(X_{\Delta n},\, n\in\mathbb{N})$ is

$$L_n(\zeta)=\prod_{i=1}^n p_X\left(\Delta,X_{i\Delta}|X_{(i-1)\Delta};\zeta\right),$$

and the log-likelihood function is

$$l_n(\zeta)=\sum_{i=1}^n\ln\left(p_X\left(\Delta,X_{i\Delta}|X_{(i-1)\Delta};\zeta\right)\right).$$

The transition density $p_X\left(\Delta,X_{i\Delta}|X_{(i-1)\Delta};\zeta\right)$ is rarely known in explicit form, which is also the case for the GGDiff. However, it is known that the transition density converges to the stationary density. Therefore, for the purpose of estimation of parameter ζ we disregard the existing exponentially decaying autocorrelation structure of the diffusion and define the quasi-likelihood function simply as

$$L_n(\zeta)=\prod_{i=1}^n f_{s,b}(X_{\Delta i}),$$

where $f_{s,b}$ is the GGDiff stationary density given by (1) for $b=0$. Then the quasi log-likelihood function is

$$l_n(\zeta)=\sum_i^n\ln\left(\frac{1}{2(s\sigma^2)^{1/s}\Gamma\left(1+\frac{1}{s}\right)}e^{-\frac{|X_i|^s}{s\sigma^2}}\right), \tag{13}$$

with the obvious simplification $\Delta = 1$ in the construction of the time-series of observations. The estimate $\hat{\zeta} = (\hat{s}, \hat{\sigma^2})$ of the parameter $\zeta = (s, \sigma^2)$ is then obtained by maximising (13), which can be performed using existing non-linear optimization methods. For more details on maximum likelihood estimation for diffusion processes we refer to [4] and [8].

4.2 Tail index estimation

Recall that for $b > 0$ the distribution (1) is heavy-tailed. The tails of this distribution decay as $|x|^{-1-s\left(\frac{\sigma^2}{b}+1\right)}$, so the tail index is of the form $\alpha = s\left(\frac{\sigma^2}{b}+1\right)$. There are many methods for tail index estimation, see e.g., [12] for a concise overview. In this paper we use the approach introduced in [13] based on the empirical scaling function. The shape of the scaling function is strongly influenced by the tail index, and graphical inspection is used for estimation of the tail index of the corresponding distribution.

Let $X_1, X_2, \ldots, X_n$ be a zero mean sample coming from a stationary heavy-tailed sequence with strong mixing property with an exponentially decaying rate. The partition function of this sample is defined by the following expression:

$$S_q(n,t) = \frac{1}{\lfloor n/t \rfloor} \sum_{i=1}^{\lfloor n/t \rfloor} \left| \sum_{j=1}^{\lfloor t \rfloor} X_{(i-1)\lfloor t \rfloor + j} \right|^q,$$

where $q > 0$ and $1 \le t \le n$. For a fixed value of q, the scaling function S_q can be estimated as follows using the empirical scaling function based on a finite sample and chosen points $s_i \in (0,1)$, $i = 1, \ldots, N$:

(14)
$$\hat{\tau}_{N,n}(q) = \frac{\sum_{i=1}^N s_i \frac{\ln S_q(n, n^{s_i})}{\ln n} - \frac{1}{N}\sum_{i=1}^N s_i \sum_{j=1}^N \frac{\ln S_q(n, n^{s_j})}{\ln n}}{\sum_{i=1}^N (s_i)^2 - \frac{1}{N}\left(\sum_{i=1}^N s_i\right)^2}.$$

By repeating this for different values of q, a plot of the empirical scaling function $\hat{\tau}_{N,n}$ is obtained.

Then estimation of tail index is based on the asymptotic behaviour of the empirical scaling function $\hat{\tau}_{N,n}$ as $n, N \to \infty$. For each $q > 0$, the limit in probability is

(15)
$$\tau_\alpha^\infty(q) = \begin{cases} \frac{q}{\alpha}, & \text{if } q \le \alpha \text{ and } \alpha \le 2, \\ 1, & \text{if } q > \alpha \text{ and } \alpha \le 2, \\ \frac{q}{2}, & \text{if } 0 < q \le \alpha \text{ and } \alpha > 2, \\ \frac{q}{2} + \frac{2(\alpha-q)^2(2\alpha+4q-3\alpha q)}{\alpha^3(2-q)^2}, & \text{if } q > \alpha \text{ and } \alpha > 2, \end{cases}$$

where α is the tail index.

Clearly, the shape of the empirical scaling function strongly depends on the value of the tail index, and the sample-based empirical counterpart $\hat{\tau}_{N,n}(q)$ can be used to estimate the index. The asymptotic shape of τ_α^∞ is shown in Figure 1. Heavy-tailed samples are characterized by the (approximately) broken line shape of the empirical scaling function, with the break occurring at α. The limiting case as $\alpha \to \infty$ corresponds to the light-tailed distributions with the straight line scaling function $q/2$ (dotted in Figure 1). Estimation can be done by fitting the empirical scaling function to its asymptotic form. For arbitrary points $q_i \in (0, q_{max})$, $i = 1, \ldots, M$, the tail index estimate is given by

(16)
$$\hat{\alpha} = \arg\min_{\alpha \in (0,\infty)} \sum_{i=1}^M (\hat{\tau}_{N,n}(q_i) - \tau_\alpha^\infty(q_i))^2.$$

Note that the advantage of this approach is that the estimation does not depend on the particular form of the underlying distribution.

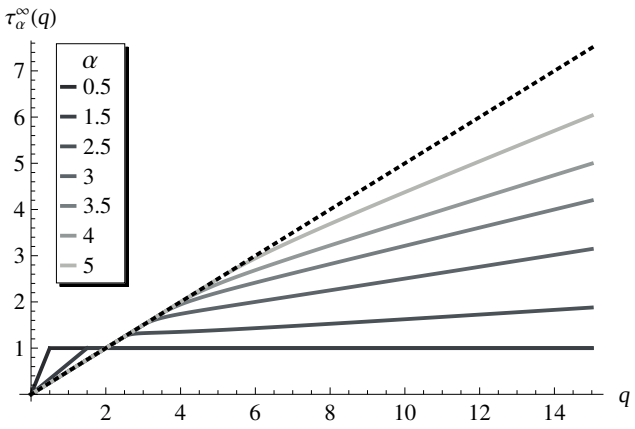


Figure 1: Asymptotic form of scaling function.

Remark 4.1. Within this paper we will use the special Student case ($s = 2$) of the $b > 0$ GGD (1), introduced in 2.1. For method of moments estimation of parameters of Student distribution we refer to [24]. Using parametrization of Student distribution (6), in order to estimate the tail index, it is enough to estimate the parameter σ^2/b. Here we provide the method of moments estimator of the parameter $\kappa = \sigma^2/b$

$$\widehat{\kappa} = \frac{m_4}{m_4 - 3},$$

based on the forth empirical moment

$$m_4 = \frac{1}{n}\sum_{k=1}^n X_k^4$$

of the set of discrete observations $(X_1, \ldots, X_n)$.

5. ANALYSIS OF EEG DATA

Data used in this analysis were used previously in [40]. Data were collected during the observational study of the pathogenesis of severe malaria (cerebral malaria and severe malarial anemia) in surviving children. Data collection was performed at Mulago National Referral and Teaching Hospital in Kampala, Uganda between 2008 and 2015. The observational study was approved by the Institutional Review

Boards of the Makerere University School of Medicine and the University of Minnesota. Study also included community control children who were recruited from household compound area of children with cerebral malaria or severe malarial anemia to control for socioeconomic variables that affect neurodevelopment and cognition [1]. Children enrolled in the study were between 18 months and 12 years of age. As cerebral malaria results in a coma, EEG signals were recorded during coma for the children who were diagnosed with cerebral malaria.

5.1 Description of EEG data

EEG data for 78 children were recorded using the International 10–20 system with the sampling rate of 500 Hz and the average record duration of 30 minutes, which resulted in approximately 1 million data points per channel. Since the Cz electrode is placed at the centre of midline sagittal plane, it was chosen as the reference electrode. Artefacts due to breathing, muscle movement and heartbeat had already been removed from this dataset. EEG signal for every child was the result of a recording from 19 channels. Data from some channels for some children contained a substantial number of zero observations, potentially due to a poor connection between the electrode and the scalp. This consideration was included in the interpretation of the results as discussed below.

5.2 Fitting of light-tailed GGD to EEG increments

Histograms of the increments of EEG signals were investigated for every child and every channel separately resulting in a total of 1492 histograms. Histograms were approximately symmetrical with means close to zero but also displayed higher peaks than the normal distribution, which further justifies the choice of GGD (1) as the marginal distribution for modeling. Estimation of parameter $\zeta = (s, \sigma^2)$ of light-tailed GGD(s, σ^2) was obtained using quasi-likelihood approach presented in 4.1. Non-linear optimization was performed using `maxLik` package [15] in `R` version 4.0.4 for Windows. Due to the fact that the shape parameter s appears inside the Γ-function in (1), constrains for the minimum value of s were used in the optimization to prevent computational problems. Optimization was performed on the entire dataset of EEG increments values and no sampling was used.

Values of $\hat{s}$ were in the range of $[0.02, 8.23]$. Extremely low values of $\hat{s}$ (close to 0.02) and $\widehat{\sigma^2}$ (close to 0.1) appeared on the channels containing very large number of zero observations and manifested in histograms of unusually high and narrow peaks (2a). Values of $\hat{s}$ near 2 result in a GGD fit that resembles a zero-mean normal distribution, which can be seen in Figure 2b. Figure 2c and Figure 2d show examples of histograms, where obtained estimates for the shape parameter were $\hat{s} < 2$ and $\hat{s} > 2$, respectively. Dashed line in the plots represents the fitted GGD.

5.3 Estimation of tail index on EEG increments

Tail index estimation of EEG increments was performed using the graphical method presented in Section 4.2. The estimation was divided into two cases, $\alpha \leq 2$ and $\alpha > 2$, due to complexity of the expression τ_α^∞ in (15). This means that the corresponding part of τ_α^∞ based on the true value of α was used as a model function for the estimate $\hat{\alpha}$ in (16). Hence, before obtaining the value of the estimate $\hat{\alpha}$, it was necessary to visually inspect the plot of empirical scaling function and determine where the break point happens.

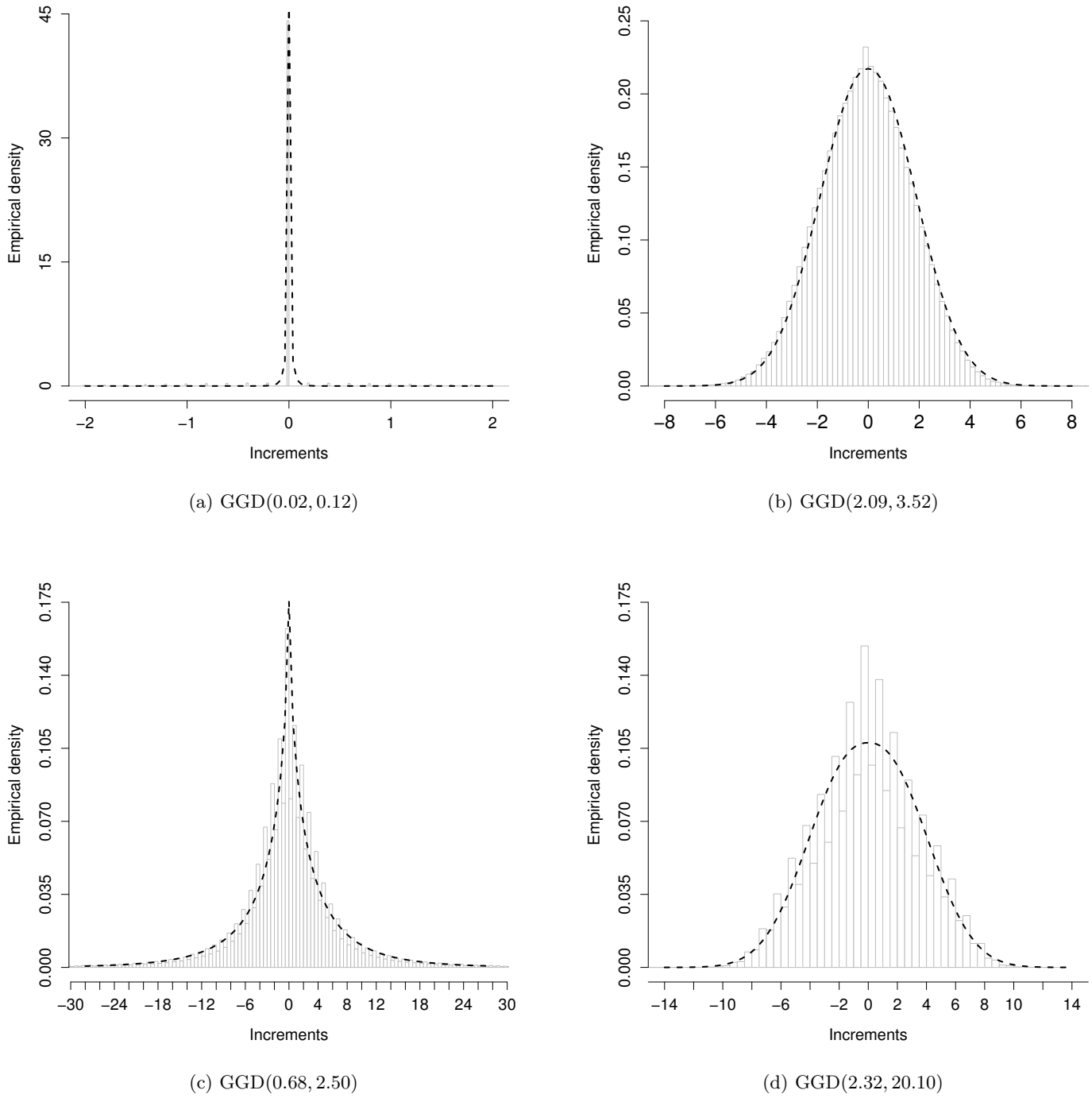
To obtain the numerical value of $\hat{\alpha}$, the empirical scaling function was fitted to the asymptotic form τ_α^∞ using ordinary least squares method. For the calculation of $\hat{\tau}_{N,n}$ values for $s_i \in (0, 1)$ were chosen to be equidistant points $(N = 23)$ in the interval $[0.1, 0.9]$, while for q_j a total of 40 equidistant points were taken from interval $[0.11, 10]$. A random sample of EEG increments of size 10000 was chosen for every channel, and the sampling was repeated 10 times. Empirical scaling function was plotted for each of these 10 samples (shown by dot-dashed lines in Figure 3), obtaining estimates $\hat{\alpha}_i$, $i = 1, \ldots, 10$. The final value of tail index estimate $\hat{\alpha}$ was chosen to be the median of values $\hat{\alpha}_i$, represented by a solid line in Figure 3. Bilinear shape (i.e a broken line) is clearly visible on both plots and identifies the breakpoint which determines whether the data are from the distribution with infinite (Figure 3a) or finite (Figure3b) variance. Analysis was performed using `Mathematica` version 11.3 for Windows.

6. PREDICTION OF NEURODEVELOPMENTAL AND COGNITIVE SCORES

6.1 Measures included in the study

Age-appropriate assessments of neurodevelopment and cognition were performed using the Mullen Scales of Early Learning [28] for children 5 years of age or younger. For children over the age of five, cognitive assessments was performed using Kaufman Assessment Battery for Children, second edition [21]. In the absence of Ugandan norms, all measures were standardized using the available United States norms. A single measure of neurodevelopment and cognition regardless of age was then obtained by computing the z-scores using the mean and standard deviation of the community control children, as discussed in [40]. For all children, assessment was performed at three points in time - after the discharge from the hospital (for community control children, this was the point of enrolment), 6 months after the discharge (or enrolment) and 12 months after the discharge (or enrolment).

Other non-EEG data that were collected in the study included demographic and anthropometric characteristics (age, sex, height-for-age and weight-for-age-z-score using the



(a) GGD(0.02, 0.12)

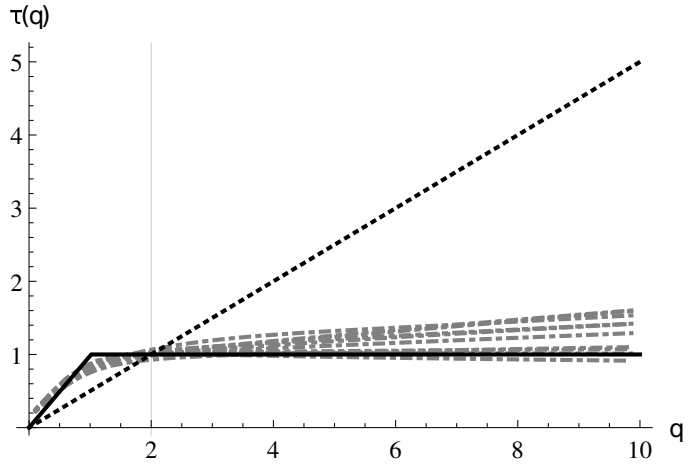
(b) GGD(2.09, 3.52)

(c) GGD(0.68, 2.50)

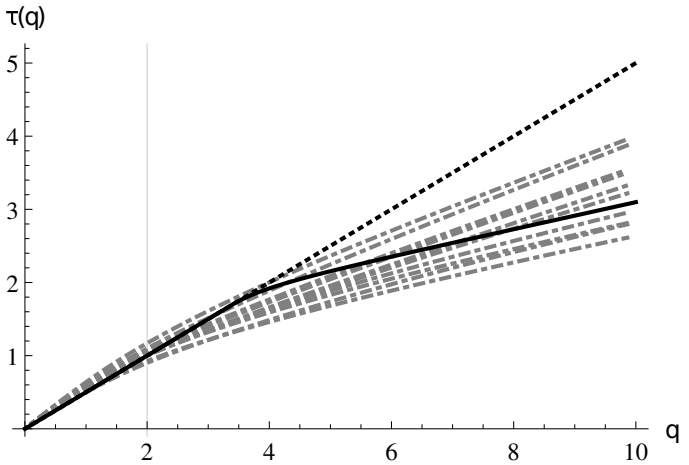
(d) GGD(2.32, 20.10)

Figure 2: Fitting of light-tailed GGD to EEG increments.

The World Health Organization reference norms [42]). Socioeconomic status was measured using a checklist of material possessions, housing quality, cooking resources and water accessibility. Quality and quantity of stimulation to which the child is exposed in the home environment was assessed using Home Observation for the Measurement of the Environment (HOME) measure [5], where higher HOME scores indicate higher quality of home environment. Severity of coma caused by cerebral malaria was assessed using the Blantyre coma scale [38]. At the point of hospitalization



(a) $\hat{\alpha} = 1.10$



(b) $\hat{\alpha} = 3.49$

Figure 3: Tail index estimates of EEG increments.

of children with cerebral malaria, biomarker panels from plasma and cerebrospinal fluid were collected. Preprocessing of the data included location and scale transformation for all potential predictors included in feature matrices.

6.2 Methods and models used

To identify important predictors of neurodevelopment and cognition 6 months after coma from cerebral malaria, elastic net regression was used. The method was introduced by [44] as a way of controlling for correlations among predictors and dealing with the case where the number of predictors is much bigger than the number of observations. Elastic net regression can be viewed as a penalized least squares method which minimizes the loss function defined by

$$L(\alpha, \lambda, \boldsymbol{\beta}) = |\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}|^2 + \lambda \left(\frac{1-\alpha}{2} |\boldsymbol{\beta}|^2 + \alpha |\boldsymbol{\beta}|_1 \right), \tag{17}$$

where

$$|\boldsymbol{\beta}|^2 = \sum_{j=1}^{p} \beta_j^2, \qquad |\boldsymbol{\beta}|_1 = \sum_{j=1}^{p} |\beta_j|$$

$\boldsymbol{y} = (y_1, \ldots, y_n)^\tau$ is the response, $\boldsymbol{X} = (\boldsymbol{x}_1 | \ldots | \boldsymbol{x})_p$ is the model matrix and $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})^\tau$, $j = 1, \ldots, p$ are the predictors [44].

Hyperparameter α can be seen as a mixing parameter between ridge ($\alpha = 0$) and LASSO ($\alpha = 1$) regression. Tuning of hyperparameters α and λ was performed using `caret` package [23] with leave-one-out cross validation. Tuning grid was constructed from values of $\lambda \in \{10^{-5}, 10^{-4}, \ldots, 10^{3}\}$ and 7 equidistant points from interval $[0.0001, 1]$ for α. Pair which had the lowest root mean squared error (RMSE) was chosen for the final model.

Since the `caret` package doesn't provide standard errors of the coefficients in the elastic net regression model, standard errors of predictors' coefficients were obtained by bootstrapping. For this purpose `R` package `boot`[7] was used with the number of bootstrap replicates set to 1000.

The response variable was the standardized neurodevelopment or cognitive score taken 6 months after the discharge from the hospital, and the scores were in the range of $[-1.99, 1.5]$. Predictor variables were taken from three sets of features. First feature matrix included just the non-EEG features described in Section 6.1, while the other two feature matrices included a combination of non-EEG features and EEG features obtained from analysis of increments. In all three feature matrices, we included socio-demographic characteristics and the neurodevelopmental or cognitive score immediately after discharge from the hospital. We refer to this score as baseline ND in reporting of the results. The rationale for the inclusion of this score was that it can be obtained relatively easily compared to the estimation of the EEG parameters for stochastic models. Fitting of stochastic models would be warranted if the EEG parameters were shown to be important over and above other measures that can be obtained easily or as part of routine clinical care for cerebral malaria.

Some features contained missing values, especially in the non-EEG dataset. Since the dataset comprised both continuous and categorical data, imputation methods suitable to mixed-type data were used. Based on the comparison of different imputation methods presented in [35], `missForest` package for `R` was selected. The algorithm is based on iterative imputation scheme by training a random forest on observed values in a first step, followed by predicting the missing values and then proceeding iteratively [37].

6.2.1 Non-EEG features model

Feature matrix for this model included a total of 54 non-EEG features with 78 observations. Maximum number of missing values per feature was 23, and most of the empty entires occured in biomarker panels from cerebrospinal fluid. Categorical variables such as *sex* (2 levels) and *bcs* (Blantyre

coma score, 6 levels) were coded into dummy variables for the inclusion in the elastic net regression. For variable *bcs*, a score of 0 (poor results) was chosen as the reference level.

6.2.2 Combined non-EEG and GGD features model

Feature matrix for this model was a combination of aforementioned 54 non-EEG features and additional 38 features (for 19 channels) which were estimates of s and σ^2 parameters obtained by fitting light-tailed GGD to EEG increments (see Section 5.2). There were no missing values in the GGD features subset.

6.2.3 Combined non-EEG and tail index features model

Feature matrix for this model was a combination of the same 54 non-EEG features and additional 19 features (for 19 channels) which were median values of estimates $\hat{\alpha}$ of tail index (described in Section 5.3). Missing values occurred for two cases and were imputed using `missForest` package. Median values of tail index estimates in the feature matrix were in the range of $[0.01, 8.76]$. To reduce the noise of this variable within the model, tail index for every channel was classified into 3 levels based on distributional tertiles. Thus, values of 3.1 and 4.5 were chosen as cut-offs. Since this resulted in creating categorical variables, they were recoded into dummy variables with values below 3.1 acting as reference level. Both versions of the model (with continuous values of tail index and categorized tail index values) were examined in the analysis.

6.3 Results and comparison

After running over a grid of different combinations of tuning parameters, best tuning parameters across all models were $\alpha = 0.833$ and $\lambda = 0.1$. These values produced the lowest RMSE. Comparison of models based on the leave-one-out cross validation RMSE is given in Table 1.

Models displayed similar RMSE values but generally, the addition of EEG features resulted in the RMSE reduction. The lowest RMSE of 0.5549 was obtained for the model containing non EEG features and categorical tail index features. This model also had the smallest number of non-zero coefficients. The list of predictors selected by each model is shown in Table 2. The table shows the values of predictors' coefficients along with their standard errors obtained by bootstrapping. Predictors that weren't included as features in models are marked by N/A and those that weren't selected by a certain model are marked by "–".

7. CONCLUSION

The EEG findings on admission to the hospital have been used to predict mortality and morbidity following illness [34]. Separately, selected plasma biomarkers have been shown to be associated with cognitive impairment in pediatric severe malaria [30]. In this analysis, we combined data on plasma and cerebrospinal fluid biomarkers together with parameters of stochastic models for the EEG data to determine their usefulness for prediction of neurodevelopment and cognition 6 month following cerebral malaria illness. Previous analysis of these data showed that stochastic modelling of EEG features can better explain the variation in neurodevelopmental and cognitive outcomes of children who were affected by cerebral malaria [40]. That previous analysis was based on EEG signals split into frequency bands and employed a different stochastic model. In this paper we used the entire EEG signal to further investigate the effect of EEG features on neurodevelopment and cognition using a strictly stationary time series model with marginal generalized Gaussian distribution and exponentially decaying autocorrelation function. The GGD parametrization used in this paper was chosen because it comprised both light and heavy-tailed distributions and can address the case of peaks higher than the normal distribution. In the first stochastic model, light-tailed GGD subfamily was fitted to EEG increments to create the features for prediction. Second stochastic model was based on a heavy-tailed distribution with tail index estimated using the empirical scaling function. Additionally, model containing non-EEG features (such as sociodemographic and anthropometric data and biomarker panels) was examined to see whether including the information from EEG signals can help in explaining of variation in neurodevelopmental and cognitive scores 6 months post-coma.

Our results show that the baseline neurodevelopmental score (taken right after coma) was the most important predictor of neurodevelopment at point 6 months after coma, which was expected. The rationale behind this comes from the fact that the baseline neurodevelopmental score is a direct measure of the outcome variable, i.e., neurodevelopmental score 6 months after coma, and thus measures the same attribute but at a different time point. Other non-EEG features retained in our model generally overlap with the non-EEG features found to be important predictors in the analysis of [40] and mostly contain biomarker panels from cerebrospinal fluid and/or plasma. Also, negative model coefficient value of some features such as white blood cell count and IL-1α receptor level in plasma are in accordance with the intuitive assumption of increased inflammatory response negatively affecting child's development. However, this model performed the worst in terms of RMSE and was improved by the addition of EEG features.

The addition of EEG features from fitting of GGD and estimation of tail index resulted in an improved RMSE for both light-tailed and heavy-tailed stochastic models. Features from the non-EEG dataset which were retained in the elastic net as the important predictors for the neurodevelopment and cognition were also kept in the combined non-EEG and GGD features model. Additional feature that was selected in this case was the GGD parameter estimate σ^2 for channel T3 (temporal electrode placed on the left side of the head). The model also had a slightly lower RMSE compared to the pure non-EEG model, meaning that the

Table 1. Model comparison based on elastic net regression results

Model features included (number of features)	RMSE	Number of non-zero coefficients	Number of non-zero coefficients from EEG features subset
Non-EEG features (54)	0.5670	12	N/A
Non-EEG (54) and GGD (38) features	0.5655	13	1
Non-EEG (54) and continuous tail index features (19)	0.5670	12	0
Non-EEG (54) and categorical tail index features (38 dummy variables)	**0.5499**	**10**	1

Table 2. Predictors selected by elastic net models

Predictor	Model coefficient (SE)			
	Non-EEG features model	Non-EEG and GGD features model	Non-EEG and continuous tail index features model	Non-EEG and categorical tail index features model
Baseline ND	4.9848×10^{-1} (8.3221×10^{-2})	5.0301×10^{-1} (8.2076×10^{-2})	4.9848×10^{-1} (8.0763×10^{-2})	4.9860×10^{-1} (8.3407×10^{-2})
Blantyre coma score	-1.1657×10^{-1} (8.3327×10^{-2})	-1.0944×10^{-1} (8.4044×10^{-2})	-1.1657×10^{-1} (8.1889×10^{-2})	-1.2040×10^{-1} (8.1088×10^{-2})
Hemoglobin level	2.1215×10^{-2} (1.6438×10^{-2})	2.3149×10^{-2} (1.7477×10^{-2})	2.1215×10^{-2} (1.7165×10^{-2})	2.3891×10^{-2} (1.7051×10^{-2})
White blood cell count	-7.2054×10^{-3} (6.3639×10^{-3})	-7.1759×10^{-3} (5.7969×10^{-3})	-7.2054×10^{-3} (6.1838×10^{-3})	-8.2922×10^{-3} (6.2979×10^{-3})
Interleukin (IL)-10 csf[1] level	6.9793×10^{-3} (5.0404×10^{-3})	7.2020×10^{-3} (4.7077×10^{-3})	6.9793×10^{-3} (4.7335×10^{-3})	7.9912×10^{-3} (4.8946×10^{-3})
Age	1.4945×10^{-3} (1.5223×10^{-2})	7.2271×10^{-4} (1.2848×10^{-2})	1.4945×10^{-3} (1.4590×10^{-2})	–
IL-1α receptor level in csf	8.9480×10^{-4} (6.0332×10^{-4})	9.6438×10^{-4} (6.3683×10^{-4})	8.9480×10^{-4} (6.1714×10^{-4})	7.6489×10^{-4} (5.4993×10^{-4})
HOME score	7.1907×10^{-4} (7.5640×10^{-3})	1.5538×10^{-4} (7.9678×10^{-3})	7.4088×10^{-3} (1.4590×10^{-2})	–
IL-6 csf level	4.4212×10^{-4} (8.5593×10^{-4})	5.0712×10^{-4} (8.4274×10^{-4})	4.4212×10^{-4} (8.2988×10^{-4})	5.8932×10^{-4} (7.4943×10^{-4})
Von Willebrand factor	3.9401×10^{-5} (3.1056×10^{-4})	4.5009×10^{-5} (3.1598×10^{-4})	3.9401×10^{-5} (3.0665×10^{-4})	–
IL-8 csf level	3.0274×10^{-5} (3.4552×10^{-5})	2.7567×10^{-5} (3.2663×10^{-5})	3.0274×10^{-5} (3.3118×10^{-5})	1.9938×10^{-5} (3.1701×10^{-5})
IL-1α receptor level in plasma	-6.1415×10^{-6} (6.6809×10^{-6})	-5.5330×10^{-6} (6.1356×10^{-6})	-6.1415×10^{-6} (6.1917×10^{-6})	-6.9700×10^{-7} (4.9735×10^{-6})
T3_CZ σ^2 estimate	N/A	2.5271×10^{-8} 1.3371×10^{-8}	N/A	N/A
T6_CZ2	N/A	N/A	N/A	-1.6478×10^{-1} 1.0155×10^{-1}

[1] cerebrospinal fluid

addition of stochastic features can improve the prediction of neurodevelopment and cognition.

Further improvement in the explanation of variation in neurodevelopment and cognition was achieved by introducing tail index estimates as features into the elastic net model. Tail index treated as continuous failed to improve the model, and none of the EEG features were chosen by the model to be significant. However, categorization of tail index based on distributional tertiles brought improvements and resulted in the model with the lowest RMSE. This means there is a threshold for the effect of the tail index value on neurodevelopmental or cognitive score. EEG feature retained in the predictive model was the tail index estimate on channel T6 (temporal electrode placed on the right side of the head) as the dummy variable with level 2. Since level 1 (values of tail index lower than 3.1) was chosen as the reference level, the interpretation behind it is that tail index value above 3.1 and below 4.5 on channel T6 has an increased negative influence on neurodevelopment and cognition compared to the tail index values of less than 3.1 on the same channel. The feature manifested in the 2$^{\text{nd}}$ highest ranked value of the coefficient in the final model which could indicate it's importance over majority of other non-EEG features.

In summary, the addition of stochastic EEG modelling improved the prediction of children's brain function 6 months following coma. Further improvement can be made

by investigating other marginal distributions appropriate for modelling of EEG signal increments (e.g., the family of generalized Pearson distributions considered in [9] and, for a moderate fraction of channels, the multimodal distribution considered in [6]). Time series models similar to the one introduced in this paper can be investigated in other diseases that affect the brain and in electrical activity of other types of cells such as cells impacted by cancer and its treatment.

ACKNOWLEDGEMENTS

The authors wish to thank three referees for the comments and suggestions, which have led to the improvement of the manuscript.

Z. Salinger was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Partnership (project reference 2275322).

REFERENCES

[1] Bangirana, P., Opoka, R. O., Boivin, M. J., Idro, R., Hodges, J. S. and John, C. C. (2016). Neurocognitive domains affected by cerebral malaria and severe malarial anemia in children. *Learn. Individ. Differ.* **46** 38–44.

[2] Bibby, B. M., Skovgaard, I. M. and Sørensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* **11** 191–220.

[3] Birbeck, G. L., Molyneux, M. E., Kaplan, P. W., Seydel, K. B., Chimalizeni, Y. F., Kawaza, K. and Taylor, T. E. (2010). Blantyre Malaria Project Epilepsy Study (BMPES) of neurological outcomes in retinopathy-positive paediatric cerebral malaria survivors: A prospective cohort study. *Lancet Neurol.* **9** 1173–1181.

[4] Bishwal, J. P. N. (2007). *Parameter Estimation in Stochastic Differential Equations*. Springer-Verlag, Berlin, Heidelberg.

[5] Bradley, R. H. and Caldwell, B. M. (1979). Home observation for measurement of the environment.

[6] Cammarota, V., Marinucci, D. and Wigman, I. (2016). On the distribution of the critical values of random spherical harmonics. *J. Geom. Anal.* **26** 3252–3324.

[7] Canty, A. and Ripley, B. D. (2021). boot: Bootstrap R (S-Plus) Functions R package version 1.3-27.

[8] Chenxu, L. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *Ann. Stat.* **41** 1350–1380.

[9] Cobb, L., Koppstein, P. and Chen, N. H. (1983). Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family. *Journal of the American Statistical Association* **78** 124-130.

[10] Doukhan, P. (1994). *Mixing - Properties and Examples*. Springer.

[11] Dytso, A., Bustin, R. and Poor, H. V. (2018). Analytical properties of generalized Gaussian distributions. *J. Stat. Distrib. Appl.* **5** 2195–5832.

[12] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events: for Insurance and Finance* **33**. Springer-Verlag, Berlin, Heidelberg, Berlin.

[13] Grahovac, D., Jia, M., Leonenko, N. and Taufer, E. (2015). Asymptotic properties of the partition function and applications in tail index inference of heavy-tailed data. *Statistics (Ber).* **49** 1221–1242.

[14] Grigelionis, B. (2013). *Student's t-Distribution and Related Stochastic Processes*. Springer-Verlag, Berlin, Heidelberg.

[15] Henningsen, A. and Toomet, O. (2011). MaxLik: A package for maximum likelihood estimation in R. *Comput. Stat.* **26** 443–458.

[16] Heyde, C. C. and Leonenko, N. N. (2005). Student processes. *Adv. Appl. Probab.* **37** 342–365.

[17] Iacus, S. M. (2009). *Simulation and Inference for Stochastic Differential Equations: with R Examples* **1**. Springer.

[18] Idro, R., Marsh, K., John, C. C. and Newton, C. R. J. (2011). Cerebral Malaria: Mechanisms Of Brain Injury And Strategies For Improved Neuro-Cognitive Outcome. *Pediatr. Res.* **68** 267–274.

[19] Johnson, O. and Vignat, C. (2007). Some results concerning maximum Rényi entropy distributions. *Ann. l'Institut Henri Poincare Probab. Stat.* **43** 339–351.

[20] Jurek, Z. J. (2001). Remarks on the selfdecomposability and new examples. *Demonstr. Math.* **34** 241–250.

[21] Kaufman, A. S. (2004). *Manual for the Kaufman Assessment Battery for Children*. Circle Pines, MN: AGS Publishing.

[22] Kloeden, P. E. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, Heidelberg.

[23] Kuhn, M. (2020). caret: Classification and Regression Training.

[24] Leonenko, N. N. and Šuvak, N. (2010). Statistical inference for student diffusion process. *Stoch. Anal. Appl.* **28** 972–1002.

[25] Lutwak, E., Yang, D. and Zhang, G. (2004). Moment-entropy inequalities. *Ann. Probab.* **32** 757–774.

[26] Lutwak, E., Yang, D. and Zhang, G. (2007). Moment-entropy inequalities for a random vector. *IEEE Transactions on Information Theory* **53** 1603-1607.

[27] Mandal, S., Sarkar, R. R. and Sinha, S. (2011). Mathematical models of malaria - a review. *Malaria Journal (part of the Springer Nature)* **10** 124-130.

[28] Mullen, E. M. (1995). *Mullen Scales of Early Learning*. American Guidance Service, Circle Pines.

[29] Nadarajah, S. (2005). A generalized normal distribution. *J. Appl. Stat.* **32** 685–694.

[30] Ouma, B. J., Ssenkusu, J. M., Shabani, E., Datta, D., Opoka, R. O., Idro, R., Bangirana, P., Park, G., Joloba, M. L., Kain, K. C., John, C. C. and Conroy, A. L. (2020). Endothelial activation, acute kidney injury, and cognitive impairment in pediatric severe malaria. *Crit. Care Med.* **48** 734–743.

[31] Ozaki, T. (1985). 2 Non-linear time series models and dynamical systems. In *Time Ser. Time Domain. Handbook of Statistics* **5** 25–83. Elsevier.

[32] Patel, A. A., Jannati, A., Dhamne, S. C., Sapuwa, M., Kalanga, E., Mazumdar, M., Birbeck, G. L. and Rotenberg, A. (2020). EEG markers predictive of epilepsy risk in pediatric cerebral malaria – A feasibility study. *Epilepsy Behav.* **113**.

[33] Piryatinska, A., Terdik, G., Woyczynski, W. A., Loparo, K. A., Scher, M. S. and Zlotnik, A. (2009). Automated detection of neonate EEG sleep stages. *Comput. Methods Programs Biomed.* **95** 31–46.

[34] Postels, D. G., Wu, X., Li, C., Kaplan, P. W., Seydel, K. B., Taylor, T. E., Kousa, Y. A., Idro, R., Opoka, R., John, C. C. and Birbeck, G. L. (2018). Admission EEG findings in diverse paediatric cerebral malaria populations predict outcomes. *Malar. J.* **17** 208.

[35] Robin, G., Mayer, I. and Sportisse, A. (2021). How to impute missing values?

[36] Sanei, S. (2013). *Adaptive Processing of Brain Signals*. John Wiley & Sons, Ltd.

[37] Stekhoven, D. J. and Buhlmann, P. (2012). MissForest–nonparametric missing value imputation for mixed-type data. *Bioinformatics* **28** 112–118.

[38] Taylor, T. E. (2009). Caring for children with cerebral malaria: insights gleaned from 20 years on a research ward in Malawi. *Trans. R. Soc. Trop. Med. Hyg.* **103** S6–S10.

[39] R Core Team (2021). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.

[40] Veretennikova, M. A., Sikorskii, A. and Boivin, M. J. (2018). Parameters of stochastic models for electroencephalogram data as

biomarkers for child's neurodevelopment after cerebral malaria. *J. Stat. Distrib. Appl.* **5**.
[41] Wanduku, D. (2019). The stochastic extinction and stability conditions for nonlinear malaria epidemics. *Mathematical Biosciences and Engineering* **16** 3771-3806.
[42] WHO (2009). Wold Health Organization Growth Standards. https://www.who.int/growthref/en/.
[43] Wolfram Research, Inc. Mathematica, Version 11.3. Champaign, IL, 2018.
[44] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **67** 301–320.

Nikolai N. Leonenko
School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF244AG, United Kingdom
E-mail address: LeonenkoN@Cardiff.ac.uk

Željka Salinger,
School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF244AG, United Kingdom
E-mail address: SalingerZ@cardiff.ac.uk

Alla Sikorskii,
Departments of Psychiatry and Statistics and Probability, Michigan State University, 909 Fee Road, East Lansing, MI 48824, United States of America
E-mail address: sikorska@msu.edu

Nenad Šuvak
Department of Mathematics, J.J. Strossmayer University of Osijek, Trg Ljudevita Gaja 6, HR-31 000 Osijek, Croatia
E-mail address: nsuvak@mathos.hr

Michael Boivin
Departments of Psychiatry and Neurology and Ophtalmology, Michigan State University, 909 Fee Road, East Lansing, MI 48824, United States of America
E-mail address: Boivin@msu.edu