**ORIGINAL PAPER**

# Multimodal diffusion model for increments of electroencephalogram data

N. N. Leonenko[1] · Z. Salinger[1] · A. Sikorskii[2] · N. Šuvak[3] · M. J. Boivin[4]

## Abstract

We propose a new strictly stationary strong mixing diffusion model with marginal multimodal (three-peak) distribution and exponentially decaying autocorrelation function for modeling of increments of electroencephalogram data collected from Ugandan children during coma from cerebral malaria. We treat the increments as discrete-time observations and construct a diffusion process where the stationary distribution is viewed as a mixture of three non-central generalized Gaussian distributions and we state some important properties related to the moments of this mixture. We estimate the distribution parameters using the expectation-maximization algorithm, where the added shape parameter is estimated using the higher order statistics approach based on an analytical relationship between the shape parameter and kurtosis. The derived estimates are then used for prediction of subsequent neurodevelopment and cognition of cerebral malaria survivors using the elastic net regression. We compare different predictive models and determine whether additional information obtained from multimodality of the marginal distributions can be used to improve the prediction.

**Keywords** Diffusion process · Multimodal distribution · Generalized Gaussian mixture · Diffusion discretization

**Mathematics Subject Classification** 37M10 · 62M10 · 62G07 · 62J20 · 62P10

✉ N. N. Leonenko
LeonenkoN@cardiff.ac.uk

Z. Salinger
SalingerZ@cardiff.ac.uk

A. Sikorskii
sikorska@msu.edu

N. Šuvak
nsuvak@mathos.hr

M. J. Boivin
Boivin@msu.edu

1 School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF244AG, UK

2 Departments of Psychiatry and Statistics and Probability, Michigan State University, 909 Fee Road, East Lansing, MI 48824, USA

3 Department of Mathematics, J.J. Strossmayer University of Osijek, Trg Ljudevita Gaja 6, HR-31 000 Osijek, Croatia

4 Departments of Psychiatry and Neurology and Ophtalmology, Michigan State University, 909 Fee Road, East Lansing, MI 48824, USA

## 1 Introduction

Multimodal distributions arise in data from many diverse fields ranging from cosmology and astrophysics (Cammarota et al. 2014) to behavioral and neuroimaging (Sun et al. 2019). If these data represent realizations of a stochastic process with a stationary distribution, models of such processes with stationary multimodal distributions may improve upon other models that do not consider multimodality. Stationarity and distributional properties of EGG data have substantial variability in the literature. Some authors treat the EEG data as stationary, whereas others observed non-stationarity in the data itself, but stationarity in the increments. These features may depend on the study population, e.g., children versus adults, subject's state (sleeping, in coma, active), type of EEG and other factors. Stationarity and distributional property may also vary across multiple EEG channels within the same sample. Veretennikova et al. (2018) and Leonenko et al. (2023) developed stochastic models for the increments of the electroencephalogram (EEG) data from Ugandan children in coma due to cerebral malaria, where the increments

of the stochastic processes had stationary Student or generalized Gaussian distributions in the Ornstein-Uhlenbeck type process or diffusion process. In these data the observed distributions of the increments were consistently symmetrical around zero across all channels, unimodal distributions in many channels with a single peak at zero, but multimodal in other channels. By approximating multimodal distributions with unimodal, useful results have been obtained in terms of parameters of the stochastic models improving prediction of subsequent neurodevelopmental and cognitive functioning of cerebral malaria survivors over what can be projected using standard clinical features. The utility of models that allow for multiple peaks in the stationary distributions is an empirical question. The first step in answering this question is constructing the corresponding stochastic processes because a priori they may not exist. 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 probability density function. Once these processes are constructed, the methods for estimating their parameters need to be considered, and then the usefulness of the parameter estimates can be tested. In the context of Ugandan cerebral malaria data, the question is whether the overall prediction of subsequent neurodevelopment or cognition can be improved, and whether additional important predictors can be identified in more refined models that account for multimodality of the marginal distributions.

The aim of this paper is stochastic modeling and prediction of evolution of increments of the EEG signal. The outcomes of the paper could have impact to public health policies for related risk assessment and be beneficial to health specialists dealing with the cerebral malaria in low and middle income developed areas that lack medical equipment.

The paper is structured as follows. In Sect. 2 we derive the parametrization of the probability density function of the distribution of increments as a mixture of three noncentral generalized Gaussian distributions (MixGGD). In the next section, we prove the existence by construction of a diffusion process with the desired stationary distribution that allows for modeling of multiple peaks in the data. We then proceed with the methods for parameter estimation for the constructed diffusion process and obtaining the estimates using the proposed method given real data. Then the derived estimates are used for prediction of subsequent neurodevelopment or cognition of survivors of cerebral malaria. We conclude with the comparison of predictive models to inform future applications of the new diffusion models in the example population in Sect. 7.

## 2 A mixture of non-central Gaussian distributions

We parametrize the probability density function (PDF) of the distribution of increments as a mixture of three noncentral generalized Gaussian distributions and we call this distribution mixed generalized Gaussian distribution (MixGGD). The probability density function (PDF) of MixGGD is of the following form:

$$f_\zeta(x) = \sum_{k=1}^{3} w_i p_k(x|\zeta) = \sum_{k=1}^{3} \frac{w_k}{2(s_k \sigma_k^2)^{1/s_k} \Gamma\left(1 + \frac{1}{s_k}\right)} e^{-\frac{|x-\mu_k|^{s_k}}{s_k \sigma_k^2}}$$

(1)

where $\zeta = (\zeta_1, \zeta_2, \zeta_3)$, $\zeta_k = (\mu_k, s_k, \sigma_k^2, w_k)$ are parameters of the distribution. Note that $\mu_k \in \mathbb{R}$, $s_k > 0$, $\sigma_k > 0$ and $0 \leq w_k \leq 1$ such that $\sum_{k=1}^{3} w_k = 1$.

For the specific values of some parameters, the MixGGD family contains the following well-known and widely applied distributions:

- For $s_1 = s_2 = s_3 = 1$ the MixGGD becomes the mixture of three Laplace distributions,
- For $s_1 = s_2 = s_3 = 2$ the MixGGD becomes the mixture of three Gaussian distributions,
- For $s_1 = 1$ and $s_2 = s_3 = 2$ the MixGGD becomes the mixture of one Laplace and two Gaussian distributions (similarly for other combinations of values 1 and 2 of parameters $s_1, s_2, s_3$),
- For $s_1 = 1$ and $s_2 = s_3 > 0$ the MixGGD becomes the mixture of one Laplace and two generalized Gaussian distributions (similarly for other combinations of same values of parameters $s_1, s_2, s_3$),
- For $s_1 = 2$ and $s_2 = s_3 > 0$ the MixGGD becomes the mixture of one Gaussian and two generalized Gaussian distributions (similarly for other combinations of same values of parameters $s_1, s_2, s_3$).

Different combinations of $s_k$ and $w_k$ with fixed values of $\mu_k$ and $\sigma_k^2$ for a two- and three-component MixGGD are shown in Fig. 1. For the sake of consistency, throughout this paper, the leftmost peak is related to the first component in the mixture ($k = 1$), the middle peak to the second component ($k = 2$) and the rightmost peak to the third component ($k = 3$) of MixGGD. Observe that a three-component MixGGD becomes a two-component MixGGD when the weight of the middle peak is set to 0, i.e. $w_2 = 0$ and by setting $w_1 = w_3 = 0$ a three-component MixGGD is reduced to a (unimodal) generalized Gaussian distribution.

Further, we state some important properties related to moments of the MixGGD. All moments exist, without any restrictions on parameter values. The integer moment of order $n \in \mathbb{N}$ is given by
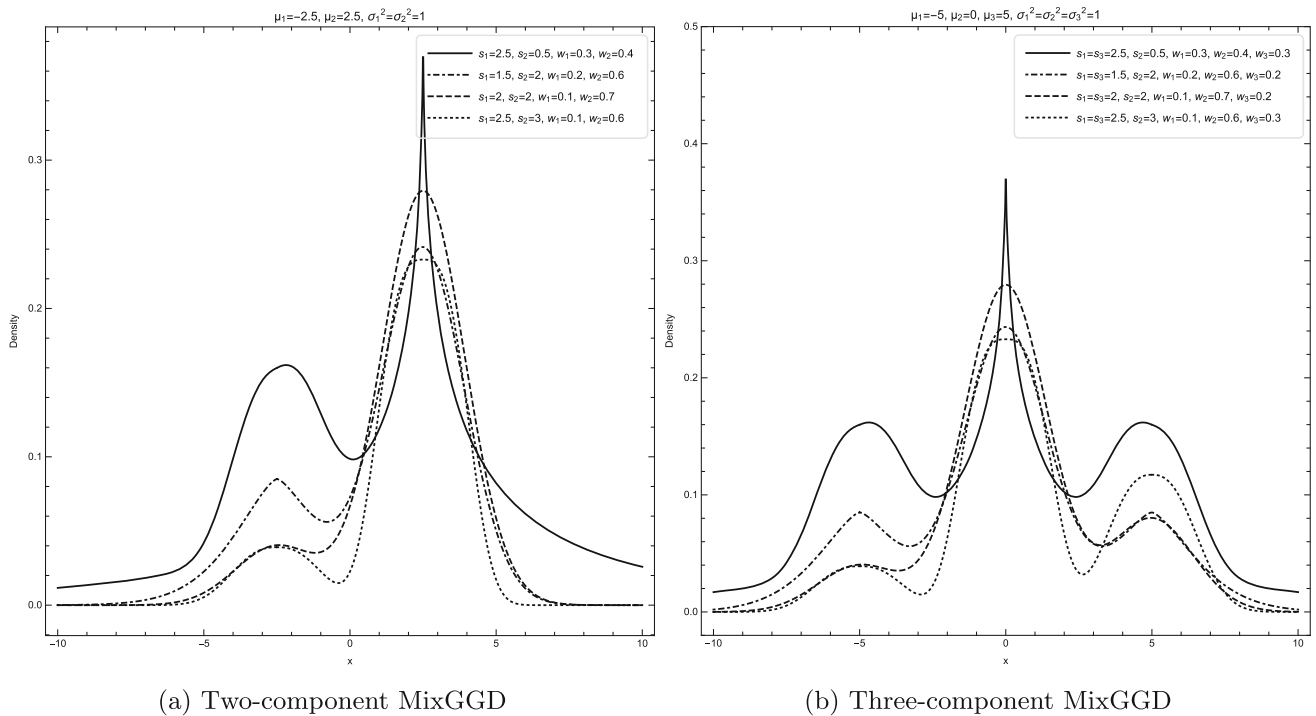
(a) Two-component MixGGD

(b) Three-component MixGGD

**Fig. 1** MixGGD with different combinations of $s_k$ and $w_k$ parameters for $K = 2$ and $K = 3$

$$E[X^n] = \sum_{k=1}^{3} w_k \mu_k^n \left( \sum_{i=0}^{n} \binom{n}{i} \left( \frac{(s_k \sigma_k^2)^{1/s_k}}{\mu_k} \right)^i \frac{(1 + (-1)^i) \Gamma\left(\frac{i+1}{s_k}\right)}{2 \Gamma\left(\frac{1}{s_k}\right)} \right),$$ (2)

while the central moments are given by

$$E[(X - E[X])^n]$$
$$= \sum_{k=1}^{3} w_k \left( \sum_{i=0}^{n} \binom{n}{i} \left( \mu_k - \sum_{k=1}^{3} w_k \mu_k \right)^{n-i} \frac{(s_k \sigma_k^2)^{i/s_k} (1 + (-1)^i) \Gamma\left(\frac{i+1}{s_k}\right)}{2 \Gamma\left(\frac{1}{s_k}\right)} \right).$$ (3)

Since the estimation of the shape parameter $s_k$ will be based on the value of kurtosis for each of the components of MixGGD (see Sect. 4), here for completeness we provide the expression for kurtosis for the MixGGD.

$$\kappa = \frac{\sum_{k=1}^{3} w_k \left( \left( \mu_k - \sum_{i=1}^{3} w_i \mu_i \right)^3 + 6 \left( \mu_k - \sum_{i=1}^{3} w_i \mu_i \right)^2 \frac{(s_k \sigma_k^2)^{2/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} + \frac{(s_k \sigma_k^2)^{4/s_k} \Gamma\left(\frac{5}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right)}{\left( \sum_{k=1}^{3} w_k \left( \left( \mu_k - \sum_{i=1}^{3} w_i \mu_i \right)^2 + \frac{(s_k \sigma_k^2)^{2/s_k} \Gamma\left(\frac{3}{s_k}\right)}{\Gamma\left(\frac{1}{s_k}\right)} \right) \right)^2}.$$

Some properties of a similar parametrization of MixGGD, where $(s_k \sigma_k^2)^{1/s_k}$ is treated as one parameter, and its relation to the multivariate generalized error distribution are studied in Wen et al. (2022). In particular, Wen et al.

(2022) consider the MixGGD with two components and present the explicit expressions for the distribution function, hazard rate function, characteristic function and some numerical characteristics (moments, central moments, skewness and kurtosis) as well as the study of moment estimation, maximum likelihood estimation and the ECM algorithm for estimation of parameters of this special case of MixGGD.

# 3 The three peak generalized Gaussian diffusion

Important probabilistic properties of EEG increments will be described by fitting these discrete-time observations to the continuous-time diffusion process $X = (X_t, t \geq 0)$ with the stationary PDF $f_\zeta$ given by (1).

Since the PDF (1) is continuous, bounded, strictly positive on $\mathbb{R}$, has expectation $\mu = \sum_{k=1}^{3} w_k \mu_k$ and finite variance, according to Bibby et al. (2005) [Theorem 2.1, page 193] the stochastic differential equation (SDE)

$$dX_t = -\theta(X_t - \mu)\, dt + \sqrt{v(X_t)}\, dB_t, \qquad t \geq 0,$$ (4)

has a unique Markovian weak solution $X = (X_t, t \geq 0)$ that is ergodic with the invariant density (1), where

- $(B_t, t \geq 0)$ is the driving Brownian motion independent of $X_0$,

- $\theta > 0$ is the parameter describing the speed of reversion of diffusion $X$ towards expectation $\mu$,
- diffusion coefficient $v(x)$ is given by

$$
\begin{aligned}
v(x) &= \frac{2\theta}{f_\zeta(x)} \int\limits_{-\infty}^{x} (\mu - y) f_\zeta(y)\, dy \\
&= 2\theta \left( \sum_{k=1}^{3} \frac{w_k}{c_k} e^{-\frac{|x-\mu_k|^{s_k}}{s_k \sigma_k^2}} \right)^{-1} \\
&\quad \left( \sum_{k=1}^{3} c_k \int\limits_{-\infty}^{x} (\mu - y) e^{-\frac{|y-\mu_k|^{s_k}}{s_k \sigma_k^2}}\, dy \right)
\end{aligned}
\tag{5}
$$

and it is strictly positive, i.e. $v(x) > 0$ for all $x \in \mathbb{R}$ (for details we refer to Bibby et al. 2005 [Lemma 2.2, page 194]).

Throughout this paper the diffusion $X$ given by the SDE (4) will be called the mixed generalized Gaussian diffusion (MixGGDiff). According to Bibby et al. (2005) [Theorem 2.1, page 193], some important properties of MixGGDiff are:

- The function $g(x) = f_\zeta(x) v(x)$ satisfies

$$
\int\limits_{-\infty}^{\infty} g(x)\, dx < \infty
$$

and

$$
E[X_{s+t}|X_s = x] = xe^{-\theta t} + \mu(1 - e^{-\theta t});
$$

- If $X_0 \sim f_\zeta$, MixGGDiff is a strictly stationary process with $E[X_t] = \mu$ for all $t \geq 0$;
- The autocorrelation function of MixGGDiff is given by $\mathrm{Corr}(X_{s+t}, X_s) = e^{-\theta t}$, $s, t \geq 0$.

**Remark 3.1** Other mixture distributions are also used in neurology, e.g. Gaussian mixture is used for estimating and comparing the shapes of distributions of neuroimaging data related to aging effects in brain white matter (Kim et al. 2014). Mixture diffusion processes are extensively used in finance, since due to their multimodality and heavier tails they often better describe the properties of financial data. For example, Alexander and Narayanan (2001) studies the option pricing under Gaussian mixture distributed returns. Furthermore, in Brigo and Mercurio (2002b) and Brigo and Mercurio (2002a) diffusions with mixture of log-normal densities are used for modeling of market smile phenomenon, while in Brigo (2008) the overview of general

properties of mixture diffusion SDEs under assumption of existence of strong solution and constant starting value is given.

## 4 Estimation of the parameters of the stationary MixGGDiff distribution

Due to the added shape parameter $s_k$, GGD is more flexible and can approximate a large class of statistical distributions. However, it also implies there is an additional parameter to estimate, compared to a Gaussian distribution. Thus, a $K$-component MixGGD requires the estimation of $4^*K$ parameters.

Method of estimation of $\zeta_k = (\mu_k, s_k, \sigma_k^2, w_k)$ in this part of the analysis was adopted from Mohamed and Jaïdane-Saïdane (2009) and is based on the expectation maximization (EM) algorithm first proposed by Dempster et al. (1977). EM algorithm is comprised of two steps: the step to find the expectation (E-step) and the maximization step (M-step). The algorithm maximizes the complete log-likelihood function

$$
l(\zeta) = \sum_{k=1}^{K} \sum_{i=1}^{N} h_{k,i} \ln\left( w_k p_k(x_i | \mu_k, s_k, \sigma_k^2) \right)
\tag{6}
$$

where

$$
h_{k,i} = p(k|x_i) = \frac{w_k p(x_i | \zeta_k)}{\sum_{j=1}^{K} w_j p(x_i | \zeta_j)}
$$

represents the conditional expectation of $p_k$ given the observation $x_i$, i.e. the posterior probability that $x_i$ belongs to the $k$th component.

Steps of the algorithm are as follows:

- *Initialization* of the model parameter $\zeta$;
- *Expectation (E-step)* where the conditional probability $h_{k,i}$ is calculated using

$$
h_{k,i}^{(n+1)} = \frac{w_k^{(n)} p(x_i | \mu_k^{(n)}, s_k^{(n)}, \sigma_k^{2(n)})}{\sum_{j=1}^{K} w_j^{(n)} p(x_i | \mu_j^{(n)}, s_j^{(n)}, \sigma_j^{2(n)})}
$$

and the computation of $l(\zeta)$ is calculated based on the estimate $\zeta^{(n)}$;

- *Maximization (M-step)* deals with the numerical maximization of the log-likelihood function (6) and the parameter estimate at iteration $(n + 1)$ is estimated as

$$
\widehat{\zeta}^{(n+1)} = \arg\max_{\zeta} l(\zeta^{(n)})
$$

Parameters $w_k, \mu_k, \sigma_k^2$ are estimated using a set of iterative equations

$$\widehat{w_k}^{(n+1)} = \frac{1}{N}\sum_{i=1}^{N} h_{k,i}^{(n)},$$

$$\widehat{\mu_k}^{(n+1)} = \frac{\sum_{i=1}^{N} h_{k,i}^{(n)} x_i}{\sum_{i=1}^{N} h_{k,i}^{(n)}},$$

$$\widehat{\sigma_k^2}^{(n+1)} = \frac{\sum_{i=1}^{N} h_{k,i}^{(n)}\left(x_i - \mu_k^{(n)}\right)^2}{\sum_{i=1}^{N} h_{k,i}^{(n)}}$$

For the added shape parameters $s_k$, approach based on the use of higher order statistics from Mohamed and Jaïdane-Saïdane (2009) was used. Specifically, values of kurtosis for each mixture component can be used to derive an approximation of shape parameters $s_k$. Tesei and Regazzoni (1998) define an analytical relationship between shape parameter $s_k$ and kurtosis $\kappa_k$ as

$$\kappa_k = \frac{\Gamma\left(\frac{5}{s_k}\right)\Gamma\left(\frac{1}{s_k}\right)}{\left(\Gamma\left(\frac{3}{s_k}\right)\right)^2} \tag{7}$$

It is impossible to express $s_k$ in terms of $\kappa_k$ due to the $\Gamma$ function definition, but an approximation can be found by applying least squared method on a generic second-order monotonic analytical expression

$$\kappa_k = \frac{1.865 s_k^2 + \alpha_1 s_k + \alpha_2}{s_k^2 + \delta_1 s_k + \delta_2}$$

which gives

$$\kappa_k \approx \frac{5 + 1.865(s_k + 0.12)^2}{(s_k + 0.12)^2}. \tag{8}$$

Kurtosis in $(n+1)$ iteration is estimated using an iterative equation

$$\kappa_k^{(n+1)} = \frac{\sum_{i=1}^{N} h_{k,i}^{(n)}\left(x_i - \mu_k^{(n)}\right)^2}{\left(\sigma_k^{(n)}\right)^4 \sum_{i=1}^{N} h_{k,i}^{(n)}}$$

with the same weights as for $\mu_k^{(n+1)}$ and $\sigma_k^{2(n+1)}$. Shape parameter $s_k$ in the $(n+1)$ iteration is then approximated by inverting the monotonic expression (8) resulting in

$$\widehat{s_k}^{(n+1)} \approx \sqrt{\frac{5}{\kappa_k^{(n)} - 1.865}} - 0.12 \tag{9}$$

which is defined for $\kappa_k > 1.865$ and gives a good approximation of shape parameter $s_k$ as a function of $\kappa_k$ when $\kappa_k$ lies in the range of $\langle 1.865, 30]$ (Tesei and Regazzoni 1996), which corresponds to $s_k$ range of $[0.302, \infty\rangle$. For $s$ in range $\langle 0.3, 10]$, comparison of true value of kurtosis and approximated value of kurtosis as a function of shape parameter $s$ based on 8 is shown in Fig. 2

**Remark 4.1** Another approach in estimating the shape parameters $s_k$ based on the numerical optimization of the log-likelihood function can be found in Mohamed and Jaïdane-Saïdane (2009). However, the approach is much more complex due to the system of equations being strongly nonlinear, sensitivity of initial parameter values and computational time.

## 5 Fitting a MixGGD to EEG data

To illustrate an application of the MixGGDiff to real data, we performed the analysis of EEG data collected in Uganda between 2008 and 2015 from 78 children 18 months to 12 years of age who were in coma due to cerebral malaria (Bangirana et al. 2016). EEG were recorded using the International 10–20 system with the sampling rate of 500 Hz and the average record duration of 30 min. The Cz electrode placed at the centre of midline sagittal plane was chosen as the reference electrode for the recordings from the 19 channels.

Previously, these data were used to fit stochastic models with unimodal marginal distributions of the increment process seen in many channels for many children (Veretennikova et al. 2018; Leonenko et al. 2023). The estimates of the parameters of the increment process were then evaluated as potential predictors of subsequent neurodevelopment or cognition for cerebral malaria survivors to determine their usefulness over and above socio-demographic characteristics, clinical factors, and child neurodevelopment (for children 5 years of age or younger) or cognitive score (for children over 5 years of age) at the time of hospitalization for cerebral malaria after coma. The use of parameters from stochastic models with unimodal distributions has improved the prediction, and some of these parameters were important. However, the observed distributions across a total of 1482 children-channels varied in terms of the number of peaks (Table 1). The question in the present application was if the analysis using more refined models that allow for multimodal distributions can produce additional useful findings.

Parameter estimation was performed without randomly sampling the values of increments, however, due to the sensitivity of the estimating equations, outliers had to be removed prior to the estimation. Outliers were defined as

**Fig. 2** Comparison of true value of kurtosis and approximated value of kurtosis as a function of shape parameter $s$
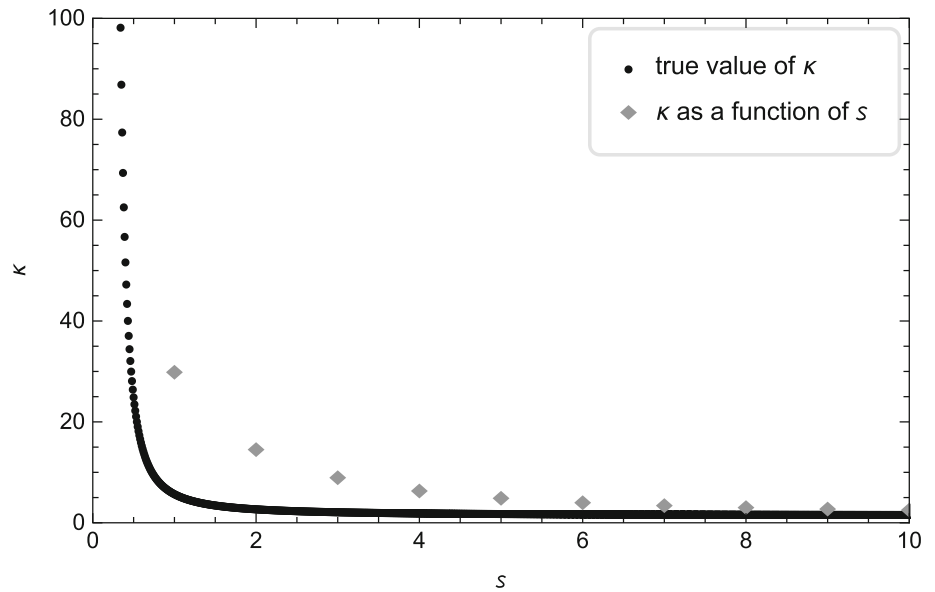
**Table 1** Classification of EEG channels based on the shape of histograms of increments

| Histogram shape | Number of channels (%) |
| --- | --- |
| 1-peak | 1100 (74.22) |
| 2-peak | 144 (9.72) |
| 3-peak | 169 (11.40) |
| 4-peak | 25 (1.69) |
| 5-peak | 40 (2.70) |
| 6-peak | 4 (0.27) |
| Total | 1482 (100) |

data points more than 1.5 of the interquartile range below the first or above the third quartile.

Although some channels' histograms displayed a distribution of up to six peaks, a GGD mixture of maximum three components ($K = 3$) was used as the model in our analysis because 95% of the person-channels had 3 or fewer peaks. Four- and six-peak channels were reclassified as having two peaks, and five-peak channels as having three peaks which can be seen as merging peaks which are in proximity of each other in the histograms.

Analysis was performed using R version 4.1.2 for Windows. EM algorithm for estimation of parameters of MixGGD wasn't available in R from any of the existing packages so the algorithm had to be implemented using the EM method described in 4. The algorithm was implemented for the case of up to three peaks and additionally, improved with the possibility of forcing a symmetrical mixture (in terms of location, scale and shape) where, for the case of two-component mixture the condition $\widehat{\mu_1} =$

$\widehat{\mu_2}$, $\widehat{\sigma_1^2} = \widehat{\sigma_2^2}$, $\widehat{s_1} = \widehat{s_2}$ was used and for the three-component mixture that condition corresponded to $\widehat{\mu_1} = \widehat{\mu_3}$, $\widehat{\mu_2} = 0$, $\widehat{\sigma_1^2} = \widehat{\sigma_3^2}$, $\widehat{s_1} = \widehat{s_3}$; only the weight parameters weren't forced to be equal.

To initialize the algorithm, $k$-means clustering available from base R package was performed, and then the initial values of parameters were calculated for each of the derived clusters. Since the channels had been previously classified based on their histogram shape, clustering was performed using the available information about the number of peaks. Initial values of parameters $\mu_k$, $\sigma_k^2$, $w_k$, along with initial value of kurtosis $\kappa_k$, for $k = 1, 2, 3$ were estimated on each of the clusters. Initial value of shape parameter $s_k$ was then calculated using the expression (9). These values of $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$ were then used as the starting points of the algorithm. Stopping criterion for the algorithm was a tolerance of $10^{-4}$ or 1000 iterations. In cases where the number of peaks wasn't conclusive, EM algorithm with all the possible number of peaks was tested and the performance compared based on the value of AIC, BIC and/or log-likelihood value. In same cases, the obtained estimates didn't present a good fit so an improvement was tried by forcing the EM algorithm for a symmetrical mixture; this resulted with an improvement in 4 channels. Algorithm converged (was considered successful) in $1263\,(85.3\%)$ of the cases (channels) and estimates couldn't be obtained in a total of 219 $(14.8\%)$ cases. Of these 219 cases where the algorithm was unsuccessful, 77 corresponded to unimodal channels with "degenerate" distribution, and the rest was on multimodal channels. The reason behind the failure of the algorithm in these cases was determined to be the value of kurtosis, which was

either lower than 1.865 or greater than 349.087 (these value of kurtosis result in negative values of shape parameter $s_k$) so the shape parameter couldn't be approximated using the expression (9).

Examples of the obtained fit for the two- and three-component MixGGD can be seen in Figs. 3 and 4, respectively. Examples of the fit where a two- or three-component MixGGD was forced on histograms with more than three peaks can be seen in Fig. 5. Values of parameter estimates are given in Table 2.

# 6 Application to prediction of neurocognitive scores

## 6.1 Methods

We investigated the MixGGD parameters as potential predictors of age-appropriate scores of neurodevelopment

and cognition at 6 months after discharge from the hospital for cerebral malaria survivors. The scores were computed as described in earlier work (Veretennikova et al. 2018; Leonenko et al. 2023) from the Mullen Scales of Early Learning (Shank 2011) for children 5 years of age or younger and Kaufman Assessment Battery for Children, second edition (Kaufman 2004) for children older than 5. The score at discharge from the hospital was used as one of the predictors, along with other predictors of neurodevelopment and cognition established in the literature: age, sex, height-for-age and weight-for-age-z-scores using the The World Health Organization reference norms (WHO 2009), socioeconomic status, quality of home environment (Bradley and Caldwell 1979), and Blantyre coma score reflecting severity of coma (Taylor 2009). In addition to these predictors, an array of standard clinical measures and biomarkers from blood and cerebrospinal fluid were included as in past work (Veretennikova et al. 2018; Leonenko et al. 2023) in order to investigate whether the
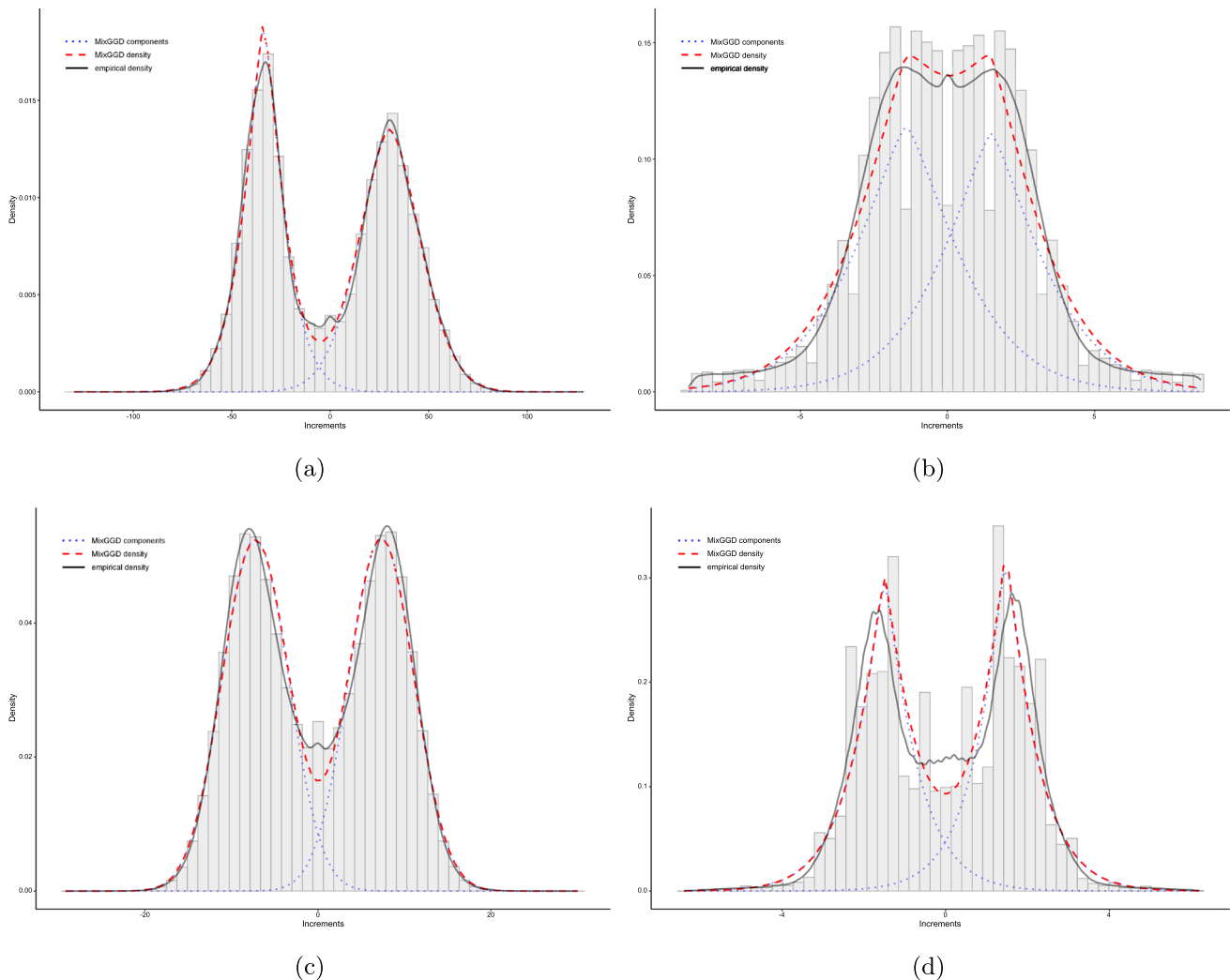


(a)



(b)



(c)



(d)

**Fig. 3** Examples of fitting a two-component MixGGD to EEG increments
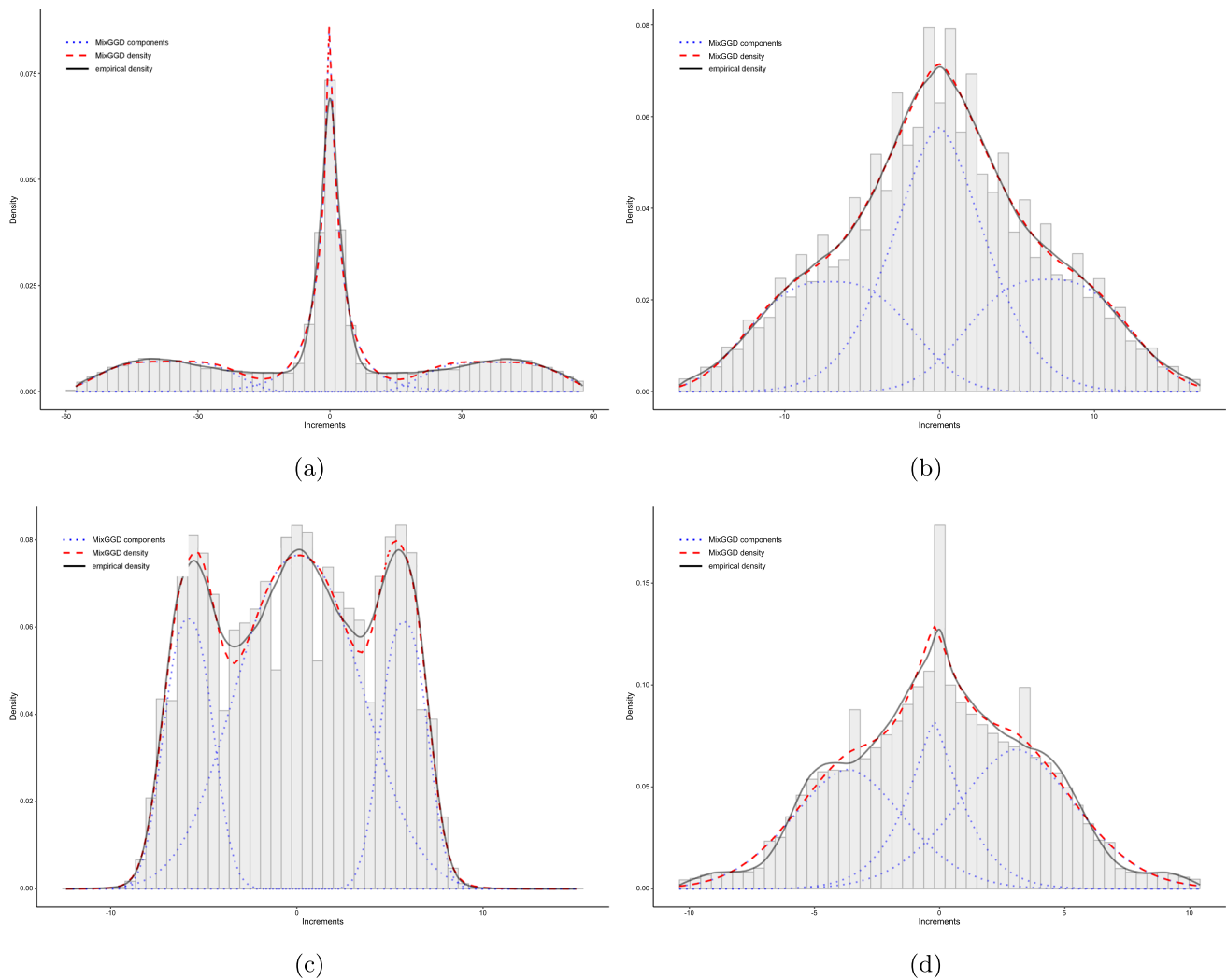
**Fig. 4** Examples of fitting a three-component MixGGD to EEG increments

EEG parameters can improve the prediction of future neurodevelopment and cognition over and above these other factors that can be obtained easily or as part of routine clinical care for cerebral malaria. The 54 non-EEG features were described elsewhere (Veretennikova et al. 2018; Leonenko et al. 2023). To these features, we added the EEG parameters reflecting multimodality. First we added the exact number of peaks (from 1 to 6) on each channel, treated as a categorical variable to reflect potentially non-linear relationship to the outcome. In the second group of models, we added the MixGGD parameters treated as continuos or categorical variables.

For the selection of predictors, we used elastic net regression. Detailed explanation of the method can be found in Zou and Hastie (2005). For our analysis, we used the same procedure explained in Leonenko et al. (2023), where the tuning of hyperparameters $\alpha$ and $\lambda$ was performed using `caret` package (Kuhn 2020) 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 $\alpha$. Pair which had the lowest root mean squared error (RMSE) was chosen for the final model.

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 $[-1.99, 1.5]$. To asses if the addition of EEG features can improve the prediction of neurodevelopment and cognition after coma, several models based on different feature matrices similar to the ones described in Leonenko et al. (2023) were constructed. 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.
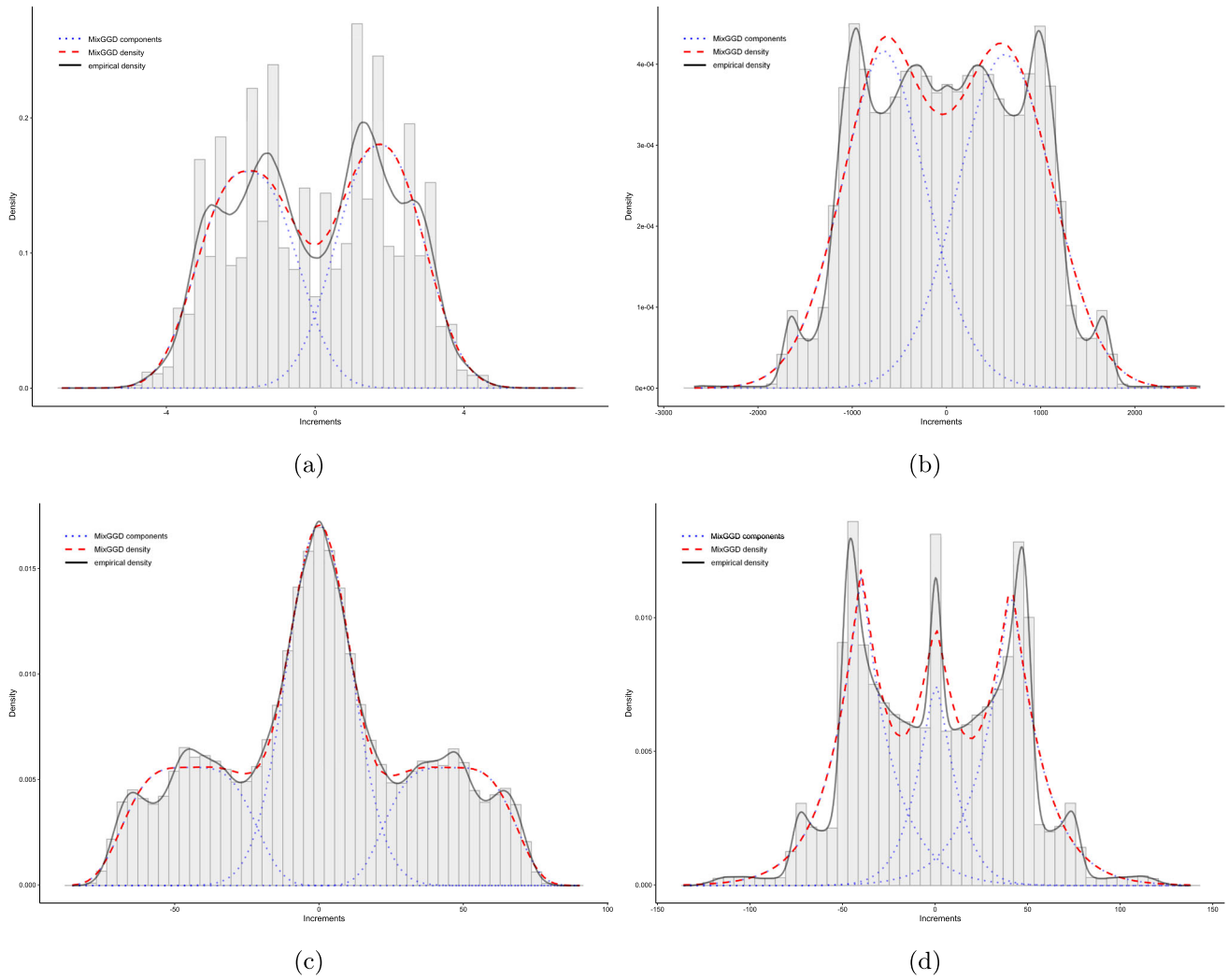
**Fig. 5** Examples of fitting a three-component MixGGD to EEG increments with more than three peaks

**Table 2** MixGGD parameter estimate values

| | MixGGD parameters | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\widehat{\mu_1}$ | $\widehat{s_1}$ | $\widehat{\sigma_1^2}$ | $\widehat{w_1}$ | $\widehat{\mu_2}$ | $\widehat{s_2}$ | $\widehat{\sigma_2^2}$ | $\widehat{w_2}$ | $\widehat{\mu_3}$ | $\widehat{s_3}$ | $\widehat{\sigma_3^2}$ | $\widehat{w_3}$ |
| Figure 3a | − 33.86 | 1.34 | 166.18 | 0.47 | | | | | 30.16 | 1.74 | 283.97 | 0.53 |
| Figure 3a | − 1.41 | 1.34 | 5.54 | 0.52 | | | | | 1.48 | 1.32 | 5.29 | 0.48 |
| Figure 3a | − 7.26 | 2.05 | 14.30 | 0.50 | | | | | 7.33 | 2.00 | 14.18 | 0.50 |
| Figure 3a | − 1.51 | 1.14 | 1.00 | 0.49 | | | | | 1.47 | 1.11 | 0.97 | 0.51 |
| Figure 4a | − 36.89 | 3.77 | 112.89 | 0.23 | 0.02 | 0.65 | 50.48 | 0.54 | 37.13 | 3.88 | 112.00 | 0.23 |
| Figure 4a | − 7.05 | 2.74 | 16.39 | 0.28 | − 0.05 | 1.67 | 11.55 | 0.44 | 6.99 | 2.70 | 16.41 | 0.28 |
| Figure 4a | − 5.82 | 2.36 | 1.32 | 0.19 | 0.11 | 2.22 | 9.69 | 0.63 | 5.74 | 2.43 | 1.18 | 0.18 |
| Figure 4a | − 3.66 | 1.80 | 5.88 | 0.34 | − 0.23 | 1.28 | 2.52 | 0.24 | 3.06 | 1.79 | 6.78 | 0.42 |
| Figure 5a | − 1.81 | 2.51 | 1.23 | 0.49 | | | | | 1.74 | 2.24 | 1.15 | 0.51 |
| Figure 5a | − 673.06 | 1.81 | 232133.93 | 0.48 | | | | | 613.24 | 1.98 | 260249.63 | 0.52 |
| Figure 5a | − 44.84 | 3.76 | 227.45 | 0.26 | − 0.11 | 2.00 | 128.94 | 0.48 | 45.27 | 3.90 | 218.94 | 0.26 |
| Figure 5a | − 40.14 | 1.07 | 502.34 | 0.40 | − 0.09 | 1.32 | 194.85 | 0.20 | 40.54 | 1.10 | 501.41 | 0.40 |

**Table 3** Model comparison based on elastic net regression results

| Model features included (number of features, including dummy) | RMSE | Number of non-zero coefficients | Number of non-zero coefficients from EEG features subset |
|---|---|---|---|
| Non-EEG features (58) | 0.5670 | 12 | N/A |
| Non-EEG + number of peaks (134) (categorical predictor with 5 levels) | 0.5530 | 17 | 2 |
| Non-EEG + MixGGD features (286) (continuos predictors) | 0.5716 | 10 | 1 |
| Non-EEG + MixGGD features (286) (all estimates (except for $\widehat{\mu_k}$) categorical with 2 levels) | 0.5693 | 21 | 13 |
| Non-EEG + MixGGD features (286) (all estimates categorical with 2 levels) | 0.5654 | 34 | 24 |

Features containing missing values from the non-EEG dataset had already been imputed for the analysis performed in Leonenko et al. (2023). The same values (hence, the same feature matrix) were used in this analysis. Missing values from the non-EEG dataset (i.e. missing values of $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$) were imputed by hand using a "random-by-peak" method. This consisted of imputing random values from successful channels, i.e. for each missing case (channel), randomly choosing one channel where the estimation was successful (with respect to the number of peaks) and using those parameter estimates in place of missing values.

### 6.1.1 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 entries occurred 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.

### 6.1.2 Combined non-EEG and number of peaks features model

Feature matrix for this model was a combination of aforementioned 54 non-EEG features and additional predictors containing the information about the number of peaks on each channel. The model contained a five-level predictor variable with the number of peaks (0 for "degenerate" channels, 1, 2, 3 or more than 3 peaks). In case of the categorical predictor variables, recoding into dummy variables was performed automatically within `caret` package (Kuhn 2020) with the first level acting as the reference level.

### 6.1.3 Combined non-EEG and MixGGD features models

Feature matrix for these models consisted of aforementioned 54 non-EEG features and parameter estimates $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$ obtained by fitting a three-component MixGGD to EEG increments. Out of these parameter estimates, first model included predictors as a continuos variable of all parameter estimates, hence for each channel a total of 12 features were included as a result of fitting a three-component MixGGD. This meant that the channels displaying one or two peaks didn't have the values of parameter estimates of some components, as these components weren't naturally present in the channel. Those values were imputed by using the value 0 both as an indicator and as a natural selection for a parameter estimate of a missing component, as this value (with the exception of location parameter $\mu_k$) wouldn't occur for any of the parameters in MixGGD. Two other models that were tested in the analysis used categorical variables with cut-off values for parameter estimates of the components. Predictors for the feature matrices of these models were formed as follows:

- Two-level for all estimates, except $\widehat{\mu_k}$, $k = 1, 2, 3$ where 0 was selected as a cut-off value for $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$, $k = 1, 2, 3$ to indicate "peak not present/present in channel"
- Two-level for all estimates, selecting 0 as a cut-off value for all $\widehat{\mu_k}$, $\widehat{\sigma_k^2}$, $\widehat{s_k}$ and $\widehat{w_k}$, $k = 1, 2, 3$ to indicate "peak not present/present in channel"

## 6.2 Results

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 3.

Models displayed similar RMSE values with the addition of categorizing the number of peaks in channels showing only negligible reduction in RMSE, but with a higher number of non-zero coefficients compared to non-EEG features model. Note that estimating the actual parameters related to multimodality did not improve upon the model where the number of peaks was accounted for without the peak parameter estimates.

# 7 Conclusion

Previous analysis of EEG data by Veretennikova et al. (2018) and Leonenko et al. (2023) showed that the addition of parameters of stochastic models from EEG data can improve the prediction of neurodevelopment and cognition over and above socio-demographic, anthropometric, and standard clinical factors. This paper builds upon the analysis of Leonenko et al. (2023) where a generalized Gaussian time series model was used to investigate the effect of EEG features on neurodevelopment and cognition. Since a detailed investigation of EEG increments displayed up to six peaks in some channels, a multimodal distribution was considered as a refinement for previously used unimodal distribution. For this purpose, a multimodal diffusion model with a mixed generalized Gaussian distribution (MixGGD) as a stationary probability density function was constructed in this paper. Then this model was applied to the EEG data to determine the extent to which modeling multimodality may improve the results obtained from stochastic models with unimodal stationary distributions. As a compromise between precision in using the exact number of peaks and keeping the number of parameters which need to be estimated within reason, MixGGD was limited to a total of three components, as there were less than 5% of channels which manifested in more than three peaks. Parameter estimation was performed using the expectation-maximization (EM) algorithm, and the added shape parameter $s$ was estimated using the approach of higher order statistics, where the value of kurtosis is used to derive an approximation of shape parameter $s$.

Several models with different sets of predictors were investigated to determine significant predictors of neurodevelopment at point 6 months after coma. Model containing non-EEG features (such as socio-demographic and anthropometric data and biomarker panels) was used as a benchmark to see whether including the information from EEG signals can help in explaining of variation in neurodevelopmental and cognitive scores 6 months post-coma. To test this, first model consisted of non-EEG features and predictors with observed number of peaks in channels as a categorical variable with five levels. Second model consisted of non-EEG features and MixGGD features as

continuos variables. Last two models consisted of non-EEG features and MixGGD features as categorical variables with a cut-off value for parameter estimates (except the location parameter) or all parameter estimates. Categorization of predictors was used to reduce the noise of the predictors and account for potential nonlinearity in their relation to the neurocognitive outcome score.

A total of ten combined (non-EEG and EEG features) models were investigated. The addition of EEG features resulted in an improvement of RMSE in seven out of ten models, but the improvement was negligible. Also, none of the models which showed an improvement had a smaller number of non-zero coefficients, which means that an improvement in RMSE is at the cost of a more complicated model, i.e. a model with a larger number of predictors. However, all of the models included at least one predictor from non-EEG dataset. The model with the lowest RMSE had the addition of number of peaks as a categorical predictor with 5 levels (0 for "degenerate" channels, 1, 2, 3 or more than 3 peaks), where two predictors from this set of predictors were chosen as significant. Models that didn't show an improvement in RMSE were the model with MixGGD features as continuos predictors, model with MixGGD features where the estimate of the shape parameter $s_k$ was a categorical predictor with 3 levels and the model with MixGGD features where estimates of $s_1$ and $s_3$ were categorical with 2 levels and estimate of $s_2$ was categorical with 3 levels.

In general, the group of models which had the addition of features relating to the number of peaks on each of the channels showed an improvement in RMSE compared both to the benchmark model and the combined non-EEG and MixGGD features models. This could mean that the information gained only from the visual investigation of EEG increments might bring an improvement in prediction of neurodevelopment, without the need for a more detailed classification of the underlying distribution and/or estimating the parameters of such distributions. However, investigating other marginal distributions appropriate for modelling of EEG signal increments should be performed to see whether an improvement in prediction can be made using a more adequate distribution.

## Declarations

## References

Alexander C, Narayanan S (2001) Option pricing with normal mixture returns: modelling excess kurtosis and uncertainty in volatility

Bangirana P, Opoka RO, Boivin MJ, Idro R, Hodges JS, John CC (2016) Neurocognitive domains affected by cerebral malaria and severe malarial anemia in children. Learn Individ Differ 46:38–44

Bibby BM, Skovgaard MI, Sørensen M (2005) Diffusion-type models with given marginal distribution and autocorrelation function. Bernoulli 11(2):191–220

Bradley RH, Caldwell BM (1979) Home observation for measurement of the environment

Brigo D (2008) The general mixture diffusion SDE and its relationship with an uncertain-volatility option model with volatility-asset decorrelation. SSRN Electron J. arXiv:0812.4052v1

Brigo D, Mercurio F (2002) Displaced and mixture diffusions for analytically-tractable smile models. Springer, Berlin, pp 151–174

Brigo D, Mercurio F (2002) Lognormal-mixture dynamics and calibration to market volatility smiles. Int J Theor Appl Finance 5:427–446

Cammarota V, Marinucci D, Wigman I (2014) On the distribution of the critical values of random spherical harmonics. J Geom Anal 26(4):3252–3324. https://doi.org/10.1007/s12220-015-9668-5

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B 39(1):1–22

Kaufman AS (2004) Manual for the Kaufman assessment battery for children. AGS Publishing, Circle Pines

Kim N, Heo M, Fleysher R, Branch CA, Lipton ML (2014) A Gaussian mixture model approach for estimating and comparing the shapes of distributions of neuroimaging data: diffusion-measured aging effects in brain white matter. Front. Public Heal. 2:32

Kuhn M (2020) caret: classification and regression training. https://cran.r-project.org/package=caret

Leonenko NN, Salinger Z, Sikorskii A, Suvak N, Boivin M (2023) Generalized Gaussian time series model for increments of EEG data. Stat Interface 16(1):17–29

Mohamed OMM, Jaïdane-Saïdane M (2009) On the parameters estimation of the Generalized Gaussian mixture model. In: Eur. Signal Process. Conf., number Eusipco, pp 2273–2277

R Core Team (2019) A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Shank L (2011) Mullen scales of early learning. Encyclopedia of clinical neuropsychology. Springer, New York, pp 1669–1671

Sun J, Li J, Zhang H (2019) Human representation of multimodal distributions as clusters of samples. PLOS Comput Biol 15(5):e1007047. https://doi.org/10.1371/journal.pcbi.1007047

Taylor TE (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(1):S6–S10. https://doi.org/10.1016/j.trstmh.2008.10.049

Tesei A, Regazzoni CS (1996) Use of fourth-order statistics for non-gaussian noise modelling: the generalized Gaussian pdf in terms of kurtosis. In: 1996 8th European signal processing conference (EUSIPCO 1996). IEEE

Tesei A, Regazzoni CS (1998) HOS-based generalized noise pdf models for signal detection optimization. Signal Process 65(2):267–281

Veretennikova MA, Sikorskii A, Boivin MJ (2018) Parameters of stochastic models for electroencephalogram data as biomarkers for child's neurodevelopment after cerebral malaria. J Stat Distrib Appl 5(1):8

Wen L, Qiu Y, Wang M, Yin J, Chen P (2022) Numerical characteristics and parameter estimation of finite mixed generalized normal distribution. Commun Stat Simul Comput 51(7):3596–3620

WHO (2009) Wold Health Organization Growth Standards. https://www.who.int/growthref/en/

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 67(2):301–320