Biomedical Signal Processing and Control 109 (2025) 107947

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Non-Gaussian modeling of sleep EEG based on a skewed scale mixture structure and its application to sleep stage analysis☆

Miyari Hatamoto [a,1], Akira Furui [a,1], Keiko Ogawa [b], Toshio Tsuji [a,*]

[a] Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-hiroshima, 739-8527, Hiroshima, Japan
[b] Graduate School of Humanities and Social Sciences, Hiroshima University, Higashi-hiroshima, 739-8527, Hiroshima, Japan

ARTICLE INFO

Dataset link: Skew Scale Mixture Model (Original data)

Keywords:
Electroencephalogram (EEG)
Sleep stage
Stochastic model
Stochastic fluctuation
Non-Gaussianity
Skew scale mixture model

ABSTRACT

Objective: Electroencephalograms (EEGs) are widely used to evaluate sleep. Changes in the shape of EEG amplitude distributions serve as useful indicators to characterize sleep stages. However, existing models lack the representational power to comprehensively capture the non-Gaussian characteristics of EEGs.
Methods: To address this limitation, we propose a novel skew-scale mixture model based on a skewed scale mixture structure. This model treats EEG amplitudes as random variables following a multivariate Gaussian distribution, whose mean vector and covariance matrix are weighted by scale and skewness parameters. These parameters are estimated using marginal likelihood maximization and used as features to quantify non-Gaussian characteristics such as tail weight and lateral asymmetry.
Results: The proposed model was validated through simulations and applied to EEG data from the Montreal Archive of Sleep Studies (MASS) dataset, which includes five sleep stages: wakefulness, REM, N1, N2, and N3. Compared to conventional probabilistic models (e.g., Gaussian and scale mixture models), the proposed model demonstrated superior ability to represent non-Gaussian characteristics, as evaluated by Bayesian Information Criterion (BIC) scores. Moreover, extracted features showed significant variation across sleep stages, reflecting stage-specific EEG characteristics such as slow waves and spindles.
Conclusion: The proposed skew-scale mixture model provides a unified framework for comprehensively representing the non-Gaussian characteristics of sleep EEGs, including lateral asymmetry.
Significance: This model offers the potential for applications such as improved classification accuracy and enhanced detection of characteristic waveforms, laying a foundation for future developments in automated sleep stage classification.

1. Introduction

Sleep is a vital physiological process that occupies approximately one-third of human life, supporting physical recovery and neurological health. According to the guidelines of the American Academy of Sleep Medicine (AASM), sleep is categorized into five stages: wakefulness (W), rapid eye movement (REM) sleep, and three non-REM sleep stages (N1, N2, N3) [1]. Defined by varying levels of brain activity and physiological changes, these stages play critical roles in memory consolidation [2] as well as in the diagnosis and management of psychiatric and sleep-related disorders [3,4].

Electroencephalograms (EEGs), recorded through scalp-attached electrodes, are commonly used for sleep stage analysis. EEG signals exhibit characteristic patterns such as slow waves, spindles, and K-complexes depending on the depth of sleep [5]. However, manual sleep stage classification by experts based on these patterns is time-consuming, requires significant expertise, and often suffers from inter-rater variability [6]. This need for reducing subjectivity and improving efficiency has driven research into the quantitative characterization of sleep EEG features.

Most feature extraction methods have focused on analyzing frequency-domain characteristics [7,8], nonlinear dynamics [9,10], or time-domain patterns [11,12]. While these approaches have proven effective, an emerging research direction involves modeling EEG amplitude distributions as random variables to explore their statistical properties. Recent studies have focused on probabilistic modeling of

☆ This work was supported in part by JSPS, Japan KAKENHI Grant Number JP20K14698.
* Corresponding author.
E-mail addresses: miyari-hatamoto@hiroshima-u.ac.jp (M. Hatamoto), akirafurui@hiroshima-u.ac.jp (A. Furui), ogawakeicom@hiroshima-u.ac.jp (K. Ogawa), tsuji-c@bsys.hiroshima-u.ac.jp (T. Tsuji).
[1] Miyari Hatamoto and Akira Furui contributed equally to this work.

https://doi.org/10.1016/j.bspc.2025.107947
Received 17 June 2024; Received in revised form 17 February 2025; Accepted 14 April 2025
Available online 5 May 2025

EEG signals, demonstrating its effectiveness in applications such as seizure detection and emotional state classification [13–15]. This approach complements conventional feature extraction methods, which primarily rely on frequency-domain characteristics, nonlinear dynamics, or time-domain patterns. By capturing statistical properties such as non-Gaussianity, these models provide additional insights that cannot be obtained through traditional analysis alone. Such an approach has the potential to enhance the understanding of sleep-related brain activity and contribute to the development of more robust EEG-based analysis frameworks.

The primary objective of this study is to formulate and validate a comprehensive framework for analyzing the non-Gaussian characteristics of sleep EEG signals. To achieve this, we propose a novel skew-scale mixture model that extends existing EEG distribution models. This extension facilitates a more comprehensive statistical analysis of EEG signals. The effectiveness of the proposed model is evaluated through simulation experiments and its application to sleep EEG data.

The remainder of this paper is organized as follows. Section 2 reviews related works, focusing on previous EEG distribution models and their applications. Section 3 describes the proposed model, parameter estimation methods, and feature extraction process. Sections 4 and 5 present the experimental setup and results, followed by discussions in Section 6. Finally, Section 7 concludes the study and outlines future directions.

2. Related works

Recent advancements in EEG analysis have demonstrated the effectiveness of modeling EEG signals as random variables to investigate their distributional properties. Such approaches complement traditional methods focusing on the time and frequency domains or nonlinear dynamics, enabling a more precise characterization of the probabilistic nature of EEG signals. Several studies have explored variations in the distributional properties of sleep EEG signals across different sleep stages, particularly in relation to sleep stage transitions [16–19]. These studies have shown that EEG signals often display heavy-tailed distributions and asymmetry, which systematically vary with sleep depth. Such findings highlight that non-Gaussianity is an inherent property of EEG signals and suggest that probabilistic modeling is essential for accurately capturing these statistical variations across different sleep stages.

To quantify the non-Gaussianity of EEG signals, Furui et al. proposed a scale mixture model [13]. By incorporating stochastic fluctuations in the covariance matrix, this model successfully represented heavy-tailed EEG distributions and achieved high accuracy in epileptic seizure detection [14]. Additionally, the scale mixture model has been applied to emotion recognition based on EEG signals, where it demonstrated superior performance in distinguishing between pleasant and unpleasant states compared to conventional feature-based methods [15]. Furthermore, this model was extended to sleep EEG analysis, showing potential for sleep stage classification by leveraging non-Gaussianity as a distinguishing feature [20]. However, despite its effectiveness in capturing the tail behavior of EEG distributions, this model is inherently symmetric and lacks the ability to represent lateral asymmetry in EEG amplitude distributions, limiting its descriptive power.

While existing studies have demonstrated the importance of non-Gaussianity in EEG analysis, they have primarily focused on either heavy tails or asymmetry, without providing a unified framework that incorporates both characteristics. Given that EEG amplitude distributions often exhibit skewness—particularly in deeper sleep stages—the inability of previous models to account for lateral asymmetry restricts their applicability. To address this limitation, this study proposes a skew-scale mixture model, which extends the conventional scale mixture model by introducing a skewing function. This modification enables the representation of both tail weight and lateral asymmetry, thereby providing a more comprehensive characterization of EEG amplitude distributions. By formulating a probabilistic model that integrates these two aspects, our approach enhances the understanding of EEG distributional properties across different sleep stages. While this study primarily focuses on model formulation and validation, the proposed framework has potential applications in automatic sleep stage classification and feature-based EEG analysis, serving as a foundation for future advancements in sleep research.

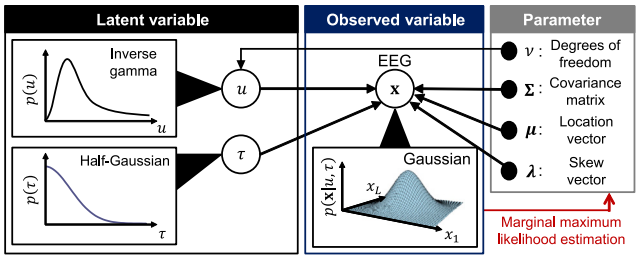

Fig. 1. Graphical representation of the stochastic relationship between the EEG signal $\mathbf{x}$ and the latent variables u and τ. The white nodes correspond to random variables, and the black nodes denote the parameters to be estimated.

3. Methods

The proposed method for analyzing sleep EEG signals involves the following key steps:

1. **Model Formulation**
   We introduce a skew-scale mixture model to capture the non-Gaussian characteristics of EEG distributions, including tail weight and lateral asymmetry.
2. **Parameter Estimation**
   The parameters of the proposed model, including the location, covariance, tail weight, and asymmetry, are estimated using an expectation–maximization (EM) algorithm.
3. **Feature Extraction**
   The estimated parameters are used to extract key features representing the distribution's location, non-Gaussianity, amplitude scale, and asymmetry for sleep EEG analysis.

3.1. Model formulation

Fig. 1 shows an overview of the proposed skew-scale mixture model. In this model, an EEG $\mathbf{x} \in \mathbb{R}^L$ measured using L electrodes is treated as a random variable following a multivariate Gaussian distribution. By weighting the covariance matrix of this distribution using a latent variable $u \in \mathbb{R}^+$ and introducing another latent variable $\tau \in \mathbb{R}^+$ related to skewness, the non-Gaussianity of the EEG amplitude distribution is represented.

3.1.1. Symmetric scale mixture model

First, we introduce the scale mixture model for EEG developed by Furui et al. [13]. In the scale mixture model, the probability distribution of an EEG $\mathbf{x}$ is given by

$$p(\mathbf{x}) = \int p(\mathbf{x}|u)p(u)\mathrm{d}u. \tag{1}$$

The conditional distribution of $\mathbf{x}$ given u is represented by the following multivariate Gaussian distribution:

$$\begin{aligned} p(\mathbf{x}|u) &= \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, u\boldsymbol{\Sigma}) \\ &= \frac{1}{(2\pi)^{\frac{L}{2}}|u\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2u}d(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\right], \end{aligned} \tag{2}$$

where $\boldsymbol{\mu} \in \mathbb{R}^L$ denotes the location vector and $\boldsymbol{\Sigma} \in \mathbb{R}^{L\times L}$ denotes the covariance matrix. $d(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the square of the Mahalanobis distance:

$$d(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = (\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}). \tag{3}$$

The latent variable u is assumed to be drawn from the following inverse gamma distribution:

$$\begin{aligned} p(u) &= \mathrm{IG}(u|\nu/2,\nu/2) \\ &= \frac{1}{\Gamma(\frac{\nu}{2})}\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} u^{-\frac{\nu}{2}-1}\exp\left(-\frac{\nu}{2}u^{-1}\right), \end{aligned} \tag{4}$$

where $\nu \in \mathbb{R}^+$ denotes a degrees-of-freedom parameter that determines the shape of the inverse gamma distribution. From (1), (2), and (4), the marginal distribution $p(\mathbf{x})$ can be expressed as

$$p(\mathbf{x}) = \int \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},u\boldsymbol{\Sigma})\mathrm{IG}(u|\nu/2,\nu/2)\mathrm{d}u. \tag{5}$$

According to (5), since the stochastic behavior of the latent variable u induces fluctuations in the covariance matrix of the Gaussian distribution, the resulting marginal distribution $p(\mathbf{x})$ is non-Gaussian with a heavier tail than the Gaussian distribution. Based on the aforementioned analysis, the EEG signal $\mathbf{x}$ is modeled using a Gaussian scale mixture distribution, i.e., an infinite mixture of Gaussian distributions with different covariance matrices [13].

3.1.2. The skewing function

We introduce the skewing function to extend the symmetric scale mixture model described in 3.1.1, enabling the model to represent the lateral asymmetry of the distribution. First, the general form of a skew-scale mixture distribution with a location vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and skew vector $\lambda \in \mathbb{R}^L$ is given by the following equation [21]:

$$p(\mathbf{x}) = 2p_0(\mathbf{x})\Phi\left[\lambda^{\mathsf{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu})\right], \tag{6}$$

where $\Phi[\cdot]$ denotes the cumulative distribution function of the standard Gaussian distribution, which operates as a skewing function containing λ to represent the lateral asymmetry of the distribution. $p_0(\mathbf{x})$ denotes an arbitrary scale mixture distribution, and by substituting (5) into it, we can derive the following skew-scale mixture model:

$$p(\mathbf{x}) = 2\int \mathcal{N}(\mathbf{x}|\boldsymbol{\mu},u\boldsymbol{\Sigma})\mathrm{IG}(u|\nu/2,\nu/2)\mathrm{d}u\Phi\left[\lambda^{\mathsf{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}-\boldsymbol{\mu})\right]. \tag{7}$$

Sampling and the derivation of the maximum likelihood estimating equation are difficult for forms involving cumulative distribution functions, such as (7). To resolve this issue, we introduce a latent variable τ that is independent of u and define the conditional distribution of $\mathbf{x}$ given u and τ as follows:

$$p(\mathbf{x}|u,\tau) = \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}+\tau\sqrt{u}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}_u, u\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\mathbf{I}_L+\lambda_u\lambda_u^{\mathsf{T}}\right)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\right), \tag{8}$$

where $\boldsymbol{\delta}_u = \lambda/\sqrt{u^{-1}+\lambda^{\mathsf{T}}\lambda}$, $\lambda_u = \sqrt{u}\lambda$ and $\mathbf{I}_L$ denotes the L-dimensional identity matrix. The latent variable τ is assumed to follow a half-Gaussian distribution:

$$\begin{aligned} p(\tau) &= \mathcal{HN}(\tau|0,1) \\ &= \sqrt{\frac{2}{\pi}}\exp\left[-\frac{\tau^2}{2}\right]. \end{aligned} \tag{9}$$

This allows the marginal distribution of $\mathbf{x}$ to be expressed as a product of probability density functions only, as follows:

$$\begin{aligned} p(\mathbf{x}) &= \iint p(\mathbf{x}|u,\tau)p(u)p(\tau)\mathrm{d}u\mathrm{d}\tau \\ &= \iint \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}+\tau\sqrt{u}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}_u, u\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\mathbf{I}_L+\lambda_u\lambda_u^{\mathsf{T}}\right)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\right) \\ &\quad \times \mathrm{IG}(u|\nu/2,\nu/2)\,\mathcal{HN}(\tau|0,1)\,\mathrm{d}u\mathrm{d}\tau. \end{aligned} \tag{10}$$

From (10), since $p(\mathbf{x})$ is marginalized with respect to the latent variables u and τ, the characteristics of the proposed model are determined by the parameters $\{\boldsymbol{\mu},\nu,\boldsymbol{\Sigma},\lambda\}$. Here, $\boldsymbol{\mu}$ determines the location; ν, the tail weight; $\boldsymbol{\Sigma}$, the spread; and λ, the lateral asymmetry. Fig. 2 illustrates examples of the change in probability density of the proposed model with respect to changes in each parameter ($L = 1$).

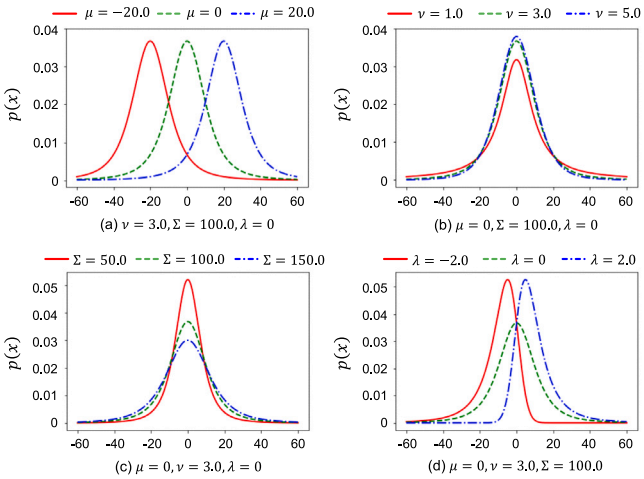

Fig. 2. Examples of changes in the probability density of the proposed model with respect to changes in the model parameters ($L = 1$). (a) Changing μ with $\nu = 3.0$, $\Sigma = 100.0$, and $\lambda = 0$. (b) Changing ν with $\mu = 0$, $\Sigma = 100.0$, and $\lambda = 0$. (c) Changing Σ with $\mu = 0$, $\nu = 3.0$, and $\lambda = 0$. (d) Changing λ with $\mu = 0$, $\nu = 3.0$ and $\Sigma = 100.0$.

In the proposed model, the non-Gaussianity of the amplitude distribution is induced by changes in the tail weight and lateral asymmetry of the distribution. Therefore, by estimating ν and λ based on the observed EEGs, the non-Gaussianity of EEGs can be evaluated in terms of two metrics: tail weight and asymmetry. In addition, by estimating $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$, the spread and location of the distribution can be evaluated.

3.2. Parameter estimation

We describe the method used to estimate the parameters of the proposed model. Given N data points of EEG signals, $\mathbf{X} = \{\mathbf{x}_n \in \mathbb{R}^L; n = 1,2,\ldots,N\}$, the parameters of the proposed model, $\{\boldsymbol{\mu},\nu,\boldsymbol{\Sigma},\lambda\}$, can be estimated by maximizing the likelihood of the marginal distribution $p(\mathbf{X}) = \prod_{n=1}^{N} p(\mathbf{x}_n)$. However, the direct optimization of $p(\mathbf{X})$ is often difficult because the maximum likelihood solution of the marginal likelihood function is generally of a complex form [22]. Therefore, in this study, the parameters are estimated using the expectation–maximization (EM) algorithm [23], which is an efficient optimization method for models with latent variables.

First, the update equation in the EM algorithm is derived by transforming the joint distribution $p(\mathbf{x}_n,u_n,\tau_n) = p(\mathbf{x}_n|u_n,\tau_n)p(u_n)p(\tau_n)$ in (10) into the following equivalent form (see Appendix):

$$\begin{aligned} p(\mathbf{x}_n,u_n,\tau_n) &= 2\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu},u_n\boldsymbol{\Sigma})\mathrm{IG}(u_n|\nu/2,\nu/2) \\ &\quad \times \mathcal{N}\left(\tau_n|\lambda^{\mathsf{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}_n-\boldsymbol{\mu}),1\right). \end{aligned} \tag{11}$$

The EM algorithm is a method that indirectly maximizes marginal likelihood by maximizing the expectation of the log-likelihood of the complete data, instead of directly maximizing the marginal likelihood. In this model, the log-likelihood of the complete data can be expressed using the joint distribution as follows:

$$L(\mathbf{x}_n,u_n,\tau_n) = \ln\prod_{n=1}^{N} p(\mathbf{x}_n,u_n,\tau_n). \tag{12}$$

Next, we estimate the parameters, $\{\boldsymbol{\mu},\nu,\boldsymbol{\Sigma},\lambda\}$, such that the marginal likelihood is maximized based on the following procedure:

(i) Initialize each parameter using an arbitrary initial value.

(ii) E-step: Calculate the expectation of the log-likelihood of the complete data, denoted by $Q(\boldsymbol{\mu}, \nu, \boldsymbol{\Sigma}, \lambda)$:

$$\begin{aligned} Q(\boldsymbol{\mu}, \nu, \boldsymbol{\Sigma}, \lambda) &= \mathbb{E}\left[L(\mathbf{x}_n, u_n, \tau_n)\right] \\ &= N \ln 2 - \frac{N}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{n=1}^{N} \left[\lambda^{\mathsf{T}} \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{x}_n - \boldsymbol{\mu})\right]^2 \\ &\quad + \frac{N\nu}{2} \ln \frac{\nu}{2} - \left(\frac{\nu + L}{2} + 1\right) \sum_{n=1}^{N} \mathbb{E}[\ln u_n] \\ &\quad - N \ln \Gamma\left(\frac{\nu}{2}\right) - \sum_{n=1}^{N} \left[\frac{\nu + d(\mathbf{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma})}{2}\right] \mathbb{E}[u_n^{-1}] \\ &\quad + \sum_{n=1}^{N} \lambda^{\mathsf{T}} \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{x}_n - \boldsymbol{\mu}) \mathbb{E}[\tau_n] - \frac{1}{2} \sum_{n=1}^{N} \mathbb{E}[\tau_n^2]. \end{aligned} \tag{13}$$

Here, $\mathbb{E}[\ln u_n]$ and $\mathbb{E}[u_n^{-1}]$ can be obtained by calculating the posterior distribution of u_n and $p(u_n|\mathbf{x}_n)$, respectively, as follows:

$$\mathbb{E}[\ln u_n] = \ln \frac{\nu + d(\mathbf{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma})}{2} - \psi\left(\frac{\nu + L}{2}\right), \tag{14}$$

$$\mathbb{E}[u_n^{-1}] = \frac{\nu + L}{\nu + d(\mathbf{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma})}, \tag{15}$$

where $\psi(\cdot)$ denotes the di-gamma function. Similarly, $\mathbb{E}[\tau_n]$ can be obtained by calculating the posterior distribution of τ_n, $p(\tau_n|\mathbf{x}_n)$, as follows:

$$\mathbb{E}[\tau_n] = \lambda^{\mathsf{T}} \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{x}_n - \boldsymbol{\mu}) + \frac{\phi[\lambda^{\mathsf{T}} \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{x}_n - \boldsymbol{\mu})]}{\boldsymbol{\Phi}[\lambda^{\mathsf{T}} \boldsymbol{\Sigma}^{-\frac{1}{2}} (\mathbf{x}_n - \boldsymbol{\mu})]}, \tag{16}$$

where $\phi[\cdot]$ denotes the probability density function of the standard Gaussian distribution.

(iii) M-step: Update the parameters by maximizing $Q(\boldsymbol{\mu}, \nu, \boldsymbol{\Sigma}, \lambda)$. To improve the efficiency of estimation, an intermediate parameter, $\boldsymbol{\Delta} = \lambda^{\mathsf{T}} \boldsymbol{\Sigma}^{-\frac{1}{2}}$, is defined, and the new estimates, ${}^{\text{new}}\boldsymbol{\mu}$,${}^{\text{new}}\boldsymbol{\Sigma}$, and ${}^{\text{new}}\lambda$, are obtained as follows:

$${}^{\text{new}}\boldsymbol{\mu} = \left(\sum_{n=1}^{N} \mathbb{E}[u_n^{-1}] \boldsymbol{\Sigma}^{-1} + N \boldsymbol{\Delta} \boldsymbol{\Delta}^{\mathsf{T}}\right)^{-1} \times \sum_{n=1}^{N} \left(\mathbb{E}[u_n^{-1}] \boldsymbol{\Sigma}^{-1} \mathbf{x}_n - \boldsymbol{\Delta} \mathbb{E}[\tau_n] + \boldsymbol{\Delta} \boldsymbol{\Delta}^{\mathsf{T}} \mathbf{x}_n\right), \tag{17}$$

$${}^{\text{new}}\boldsymbol{\Delta} = \left(\sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^{\mathsf{T}}\right)^{-1} \times \sum_{n=1}^{N} \mathbb{E}[\tau_n](\mathbf{x}_n - \boldsymbol{\mu}), \tag{18}$$

$${}^{\text{new}}\boldsymbol{\Sigma} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}[u_n^{-1}](\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^{\mathsf{T}}, \tag{19}$$

$${}^{\text{new}}\lambda = {}^{\text{new}}\boldsymbol{\Sigma}^{\frac{1}{2}}\, {}^{\text{new}}\boldsymbol{\Delta}. \tag{20}$$

Owing to the unavailability of a closed-form equation for updating ν, we adopt a binary search to identify its numerical solution that maximizes $Q(\boldsymbol{\mu}, \nu, \boldsymbol{\Sigma}, \lambda)$:

$${}^{\text{new}}\nu = \underset{\nu}{\operatorname{argmax}}\, Q({}^{\text{new}}\boldsymbol{\mu}, \nu, {}^{\text{new}}\boldsymbol{\Sigma}, {}^{\text{new}}\lambda). \tag{21}$$

(iv) Using the updated parameters, calculate the log-marginal likelihood $\ln p(\mathbf{X}) = \sum_{n=1}^{N} \ln p(\mathbf{x}_n)$ and iterate steps (ii) and (iii) until it converges.

Using the aforementioned procedure, we can estimate the parameters, $\{\boldsymbol{\mu}, \nu, \boldsymbol{\Sigma}, \lambda\}$, based on a measured EEG, $\mathbf{x}$.

3.3. Feature extraction

To evaluate sleep EEGs, we extract the following four key features from the estimated distribution parameters.

1. Non-Gaussianity corresponding to the tail weight ($1/\nu$):
   As ν approaches infinity, the amplitude distribution converges to a Gaussian distribution, reducing the non-Gaussianity in the tail. Therefore, $1/\nu$ is considered a feature representing the non-Gaussianity corresponding to the tail weight.
2. Location of the distribution ($\overline{\boldsymbol{\mu}}$):
   To evaluate the location parameter $\boldsymbol{\mu}$, we calculate the mean of all elements, $\overline{\boldsymbol{\mu}}$, which reflects the central tendency of the distribution.
3. Amplitude scale ($\|\boldsymbol{\Sigma}\|_{\text{F}}$):
   The Frobenius norm [24] of the covariance matrix $\boldsymbol{\Sigma}$ is calculated as:
   $$\|\boldsymbol{\Sigma}\|_{\text{F}} = \sqrt{\sum_{i=1}^{L} \sum_{j=1}^{L} \sigma_{ij}^2}, \tag{22}$$
   where σ_{ij} represents the elements of $\boldsymbol{\Sigma}$. This feature characterizes the overall amplitude scale of the EEG distribution.
4. Non-Gaussianity corresponding to lateral asymmetry ($\overline{\lambda}$):
   To evaluate the skewing parameter λ, the mean of all elements, $\overline{\lambda}$, is calculated. This feature captures the lateral asymmetry of the distribution.

In summary, by extracting these features $\{\overline{\boldsymbol{\mu}}, 1/\nu, \|\boldsymbol{\Sigma}\|_{\text{F}}, \overline{\lambda}\}$, the proposed method enables simultaneous evaluation of the distribution's location, tail weight non-Gaussianity, amplitude scale, and lateral asymmetry. These features provide a comprehensive representation of the EEG distribution and its characteristics across different sleep stages.

4. Experiments

4.1. Simulation

To evaluate the accuracy of parameters estimated using the proposed model, we conducted the following simulation-based experiments. First, based on ancestral sampling [22], we generated a series of random numbers following a skew-scale mixture distribution using the following procedure:

(i) Generate a discrete series, $\{u_t; t = 1, 2, \ldots, T\}$, from random numbers following the inverse gamma distribution, $\text{IG}(\nu_0/2, \nu_0/2)$.
(ii) Generate a discrete series, $\{\tau_t; t = 1, 2, \ldots, T\}$, from random numbers following the half-Gaussian distribution, $\mathcal{HN}(0, 1)$.
(iii) Generate a discrete series, $\{\mathbf{x}_t; t = 1, 2, \ldots, T\}$, from random numbers following the Gaussian distribution:
$$\mathcal{N}\left(\boldsymbol{\mu}_0 + \tau_t \sqrt{u_t} \boldsymbol{\Sigma}_0^{\frac{1}{2}} \boldsymbol{\delta}_u, u_t \boldsymbol{\Sigma}_0^{\frac{1}{2}} (\mathbf{I}_L + \lambda_u \lambda_u^{\mathsf{T}})^{-1} \boldsymbol{\Sigma}_0^{\frac{1}{2}}\right),$$
where $\boldsymbol{\delta}_u = \lambda_0 / \sqrt{u_t^{-1} + \lambda_0^{\mathsf{T}} \lambda_0}$, $\lambda_u = \sqrt{u_t} \lambda_0$.

The generated random number series, $\{\mathbf{x}_t\}$, was regarded as a pseudo-EEG signal recorded using a sampling frequency of f_{s} Hz, and the distribution parameters were estimated using the procedure described in Section 3.2. Then, we compared the true values of the parameters, $\{\boldsymbol{\mu}_0, \nu_0, \boldsymbol{\Sigma}_0, \lambda_0\}$, with the estimated values, $\{\hat{\boldsymbol{\mu}}, \hat{\nu}, \hat{\boldsymbol{\Sigma}}, \hat{\lambda}\}$ to verify the accuracy of parameter estimation. As an index of the estimation accuracy of each parameter, the absolute percentage error was defined as $|\nu_0 - \hat{\nu}|/|\nu_0| \times 100$, $\|\boldsymbol{\Sigma}_0 - \hat{\boldsymbol{\Sigma}}\|_{\text{F}}/\|\boldsymbol{\Sigma}_0\|_{\text{F}} \times 100$, $\|\lambda_0 - \hat{\lambda}\|_2/\|\lambda_0\|_2 \times 100$, and $\|\boldsymbol{\mu}_0 - \hat{\boldsymbol{\mu}}\|_2/\|\boldsymbol{\mu}_0\|_2 \times 100$, respectively.

To examine the effects of sample size and the number of input dimensions on the estimation accuracy of each parameter, the estimation window length, W, was varied among $1, 2, 5, 10, 15, 20, 30, 50$, and 100 s, and the number of dimensions, L, was varied among $2, 4, 8$, and 16. The first W s of $\{\mathbf{x}_t\}$ were used for the estimation of each parameter. In this experiment, all elements were assigned identical values for

$\boldsymbol{\mu}_0 = \{\mu_{0i}\}$ and $\lambda_0 = \{\lambda_{0i}\}$. For $\boldsymbol{\Sigma}_0$, the off-diagonal elements were taken to be 0.5 and only the diagonal elements, σ_{0ii}, were changed. The average absolute percentage error corresponding to each parameter was calculated by changing the combination of true values 240 times in aggregate: $(\mu_{0i} = -5, 5; \nu_0 = 2, 4, 6, 8, 10; \sigma_{0ii} = 1, 10, 100, 1000; \lambda_{0i} = -3, -2, -1, 1, 2, 3)$. To evaluate the calculation cost of the proposed analysis method, the computation time required for each parameter estimation was measured simultaneously. The values of T and f_s during random number series generation were taken to be 100 s and 500 Hz, respectively. The computer used in the experiment was an AMD Ryzen 7 5800X (3.8 GHz), 32.0 GB RAM.

4.2. Sleep EEG analysis

To confirm the effectiveness of the proposed model for sleep EEG analysis, EEG signals corresponding to various sleep stages were analyzed using the publicly available Montreal Archive of Sleep Studies (MASS) dataset [25]. The MASS dataset comprises overnight polysomnography signals from 200 individuals, divided into five subsets (SS1–SS5) based on acquisition protocols. In this study, we used the SS3 subset, which includes records from 62 healthy individuals (male: 29, female: 33, mean age: 42.5 ± 18.9 years). Each record consists of 20 EEG channels, 3 EMG channels, 2 EOG channels, and 1 ECG channel. EEG signals were recorded at a sampling rate of 256 Hz and processed with a 60 Hz notch filter, a 0.30 Hz low-cut filter, and a 100 Hz high-cut filter. The records were classified into five sleep stages (Stage W, R, N1, N2, N3) by experts according to the AASM standard [1], with a page size of 30 s.

In this experiment, we focused on the six primary EEG channels (F3, F4, C3, C4, O1, O2) commonly used in the AASM guidelines (see Fig. 3). Each epoch was defined as a 30-s segment, and for each sleep stage (Stage W, R, N1, N2, N3), six epochs were randomly extracted from the entire record of each participant. To ensure relatively stable segments, only epochs where the preceding and following 30 s belonged to the same sleep stage were included. Participants for whom six epochs could not be secured for all sleep stages were excluded from the analysis. As a result, we obtained data for 43 participants, consisting of five sleep stages × six epochs each.

First, the validity of sleep EEG analysis using the proposed model was evaluated through model selection based on the Bayesian information criterion [26] (BIC). The candidate group for model selection included the proposed model, the conventional scale mixture model [13], the skew Gaussian model [27], and the Gaussian model. The parameter estimation method for the proposed model is described in detail in Section 3. The distribution parameters for the comparison models were optimized using maximum likelihood estimation, similar to the proposed model, and the corresponding log-likelihoods were computed. This consistent approach ensured that the BIC values reflected a fair comparison of model fitness while accounting for model complexity. BIC, defined as:

$$\mathrm{BIC} = -2 \ln L(\hat{\theta}) + k \ln(N_W), \tag{23}$$

where $\ln L(\hat{\theta})$ denotes the log-likelihood of the model, k denotes the number of model parameters, and N_W denotes the sample size used during estimation. BIC was chosen because it balances the fitness and simplicity of a model, penalizing more complex models to prevent overfitting, and it asymptotically selects the model that is closest to the true data-generating distribution as the sample size increases. Models with a smaller BIC are thus interpreted as having a higher probability of being true models. The models of the candidate group were fitted to all epochs, and the percentages of models with the smallest BIC at different stages were compared.

Next, the relationship between the proposed feature and each stage was investigated using group comparison among all stages. We calculated the features, $\{\overline{\boldsymbol{\mu}}, 1/\nu, \|\boldsymbol{\Sigma}\|_{\mathrm{F}}, \overline{\lambda}\}$, using the distribution parameters obtained by fitting the proposed model to each epoch. Then, to account for individual differences during comparison, we introduced a generalized linear mixed model, with the stage as a fixed effect and individual differences as random effects, and conducted pairwise comparison tests based on estimated marginal means. Holm adjustment was used to correct the p values in multiple comparisons, and the significance level was taken to be 5%. For distributions of the response variables, a Gaussian distribution was taken for $\overline{\boldsymbol{\mu}}$ and $\overline{\lambda}$, and a gamma distribution was taken for $1/\nu$ and $\|\boldsymbol{\Sigma}\|_{\mathrm{F}}$ in consideration of the non-negativity of the parameters. The analyses based on generalized linear mixed models were performed using the lme4 [28] and emmeans [29] packages in R version 4.1.1. The experiments were approved by the Research Ethics Review Board of the Graduate School of Humanities and Social Sciences at Hiroshima University (Approval number: HR-PSY-002210, HR-PSY-00211).

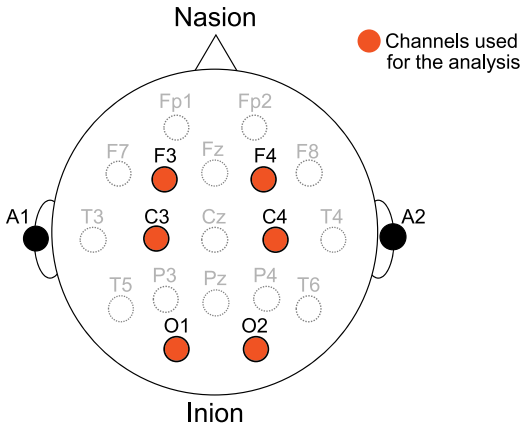

Fig. 3. International 10–20 electrode montage.

5. Results

Fig. 4 depicts examples of artificially generated pseudo-EEG signals, $\{\mathbf{x}_t\}$. In Fig. 4(a), the location of the waveform is observed to be displaced in the negative direction with $\mu_{0i} = -30.0$ and then in the positive direction with $\mu_{0i} = 30.0$. In Fig. 4(b), as ν_0 varies from $\nu_0 = 2.0$ to $\nu_0 = 10.0$, the frequency of outliers is observed to decrease, and the waveform stabilizes. In Fig. 4(c), as σ_{0ii} varies from $\sigma_{0ii} = 1.0$ to $\sigma_{0ii} = 100.0$, the amplitude of the waveform increases. In Fig. 4(d), the waveform is deflected in the negative direction with $\lambda_{0i} = -2.0$, and it is deflected in the positive direction with $\lambda_{0i} = 2.0$. Fig. 5 depicts the average absolute percentage errors for the estimation of $\{\boldsymbol{\mu}, \nu, \boldsymbol{\Sigma}, \lambda\}$ and the average computation time for different values of the window length, W, and number of dimensions, L.

Fig. 6 depicts examples of the recorded EEGs corresponding to each stage. The probability density histograms of the EEGs obtained from these data are illustrated in Fig. 7. In the figure, the results of fitting the proposed model, scale mixture model, skew Gaussian model, and Gaussian model are depicted using the solid red line, dashed blue line, dotted green line, and solid black line, respectively. In the EEGs corresponding to Stage W, R, N1 and N2, high-frequency and low-amplitude waves are dominant. In particular, in Stage N2, sudden large-amplitude waveforms that stand out from the background activity are observed (Fig. 6). In contrast, low-frequency, high-amplitude waves are observed to be dominant in Stage N3, and they are deflected in the negative direction. Table 1 lists the percentage of times that the BIC of each model is minimized in comparison with that of the other models for each sleep stage. The table also records the McNemar test results (significance level: 5%) adjusted using the Holm method by considering the proposed model as a control group. For all stages, the proposed model exhibited a significantly higher rate of achieving a minimized BIC. Furthermore, Fig. 8 presents boxplots of the BIC distributions for

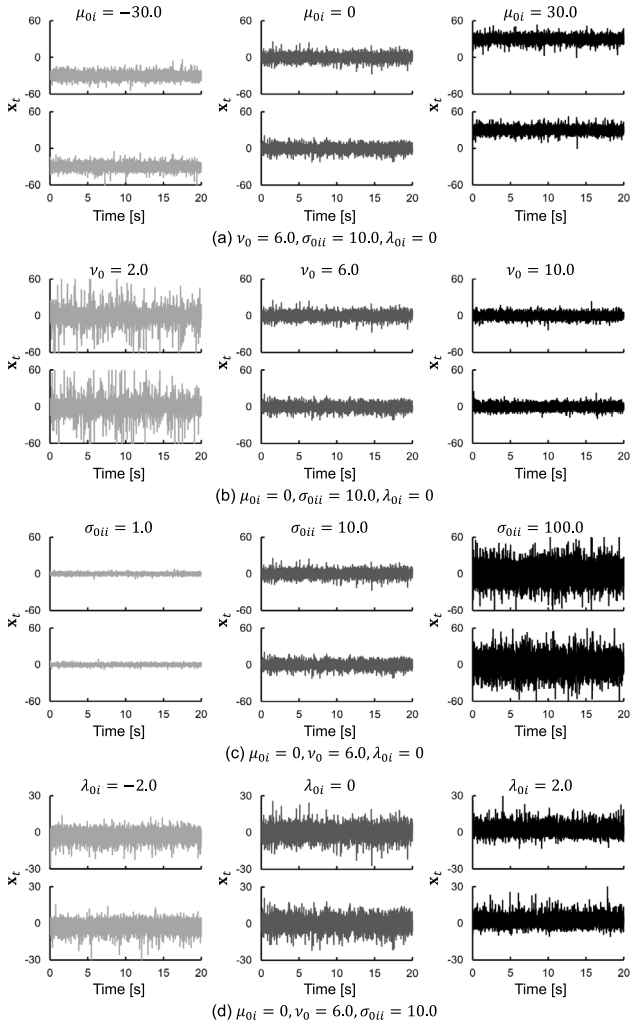

Fig. 4. Examples of generated artificial EEG signals ($W = 20$ s, $L = 2$). (a) Changing μ_{0i} with $\nu_0 = 6.0, \sigma_{0ii} = 10.0$ and $\lambda_{0i} = 0$. (b) Changing ν_0 with $\mu_{0i} = 0, \sigma_{0ii} = 10.0$ and $\lambda_{0i} = 0$. (c) Changing σ_{0ii} with $\mu_{0i} = 0, \nu_0 = 6.0$ and $\lambda_{0i} = 0$. (d) Changing λ_{0i} with $\mu_{0i} = 0, \nu_0 = 6.0$, and $\sigma_{0ii} = 10.0$.

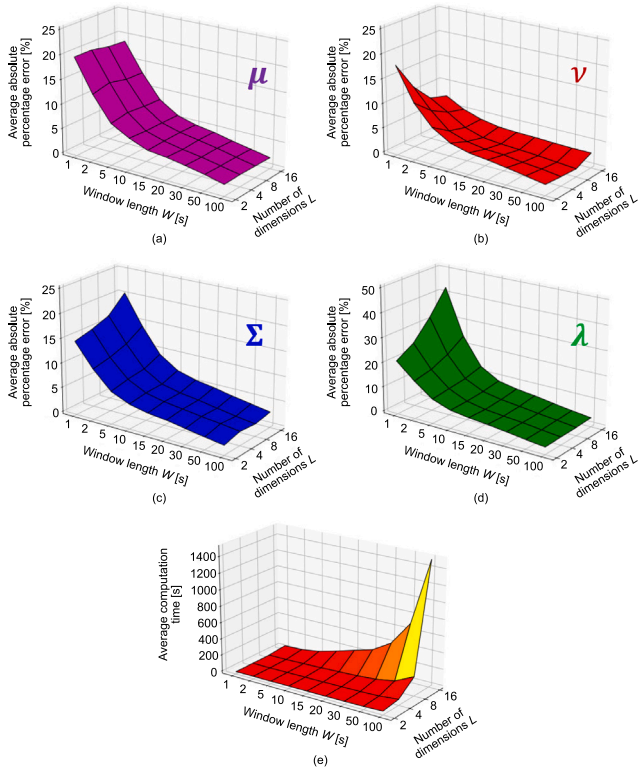

Fig. 5. Average absolute percentage errors and computation time for each combination of window length, (W), and the number of dimensions, (L), during parameter estimation. (a) Location vector, $\boldsymbol{\mu}$. (b) Degrees of freedom, ν. (c) Covariance matrix, $\boldsymbol{\Sigma}$. (d) Skew vector, $\boldsymbol{\lambda}$. (e) Average computation time.

Table 1
Percentages of times each model was selected for different sleep stages based on the BIC.

Sleep stages	Model			
	Proposed	Scale mixture	Skew Gaussian	Gaussian
W	**83.72%**	15.12%*	1.16%*	0%*
R	**89.53%**	10.08%*	0.39%*	0%*
N1	**81.40%**	17.83%*	0.77%*	0%*
N2	**82.56%**	17.44%*	0%*	0%*
N3	**95.74%**	2.71%*	1.55%*	0%*

* Significant difference with the skew-scale mixture model as identified using a McNemar test ($p < 0.001$).

each model. The figure also includes the results of the Wilcoxon signed-rank test adjusted using the Holm method. Significant differences were observed in all pairwise model comparisons. The results indicate that the proposed model consistently achieves significantly lower BIC values across various sleep stages compared to the other models.

Fig. 9 presents the box and violin plots of the calculation results of $\{\overline{\boldsymbol{\mu}}, 1/\nu, \|\boldsymbol{\Sigma}\|_F, \overline{\boldsymbol{\lambda}}\}$ for all participants at each stage. The figure also illustrates the results of pairwise comparison tests based on the estimated marginal means. Significant differences are observed in all pairwise comparisons except Stage W vs. Stage N1, Stage W vs. Stage N2, Stage R vs. Stage N1 in $\overline{\boldsymbol{\mu}}$, Stage R vs. Stage N1, Stage R vs. Stage N3 in $1/\nu$, Stage R vs. Stage N1 in $\|\boldsymbol{\Sigma}\|_F$ and Stage W vs. Stage N1, Stage W vs. Stage N2 in $\overline{\boldsymbol{\lambda}}$. Table 2 presents the estimated marginal means, with 95% confidence intervals and standard errors of the features at each stage. In $\overline{\boldsymbol{\mu}}$ and $\overline{\boldsymbol{\lambda}}$, Gaussian distributions are taken as distributions of the response variables; thus, the standard errors are pooled across stages.

6. Discussion

The simulation experiments revealed the average absolute percentage error to be approximately 19% for $\boldsymbol{\mu}$, 11% for ν, 17% for $\boldsymbol{\Sigma}$, and 30% for $\boldsymbol{\lambda}$ at $W = 1$ (Fig. 5). Meanwhile, at $W = 100$ s, the average absolute percentage error for each parameter was approximately 3%, indicating high estimation accuracy. This was attributed to the expansion of the sample size used for parameter estimation and the increasing proximity of the sample distribution and the population distribution with an increase in the estimation window. Further, as the number of dimensions, L, was increased, the average absolute percentage error decreased for ν and increased for $\boldsymbol{\Sigma}$ and $\boldsymbol{\lambda}$. This was attributed to the fact that ν is a one-dimensional parameter defined for inputs of all dimensions; thus, increasing the number of dimensions increases the sample size used for estimation and reduces the estimation error. In contrast, $\boldsymbol{\Sigma}$ and $\boldsymbol{\lambda}$ are multidimensional parameters with different estimates for each dimension. In this study, we used the Frobenius norm for $\boldsymbol{\Sigma}$ and the squared norm for $\boldsymbol{\lambda}$ to calculate the absolute percentage error. As a result, increasing the number of dimensions would have increased the overall error. In Fig. 5(e), a non-linear increase in computation time with the increase of window length W and the number of dimensions L was observed. This increase was particularly prominent when dealing with high-dimensional data, with the maximum computation time reaching approximately 1500 s when

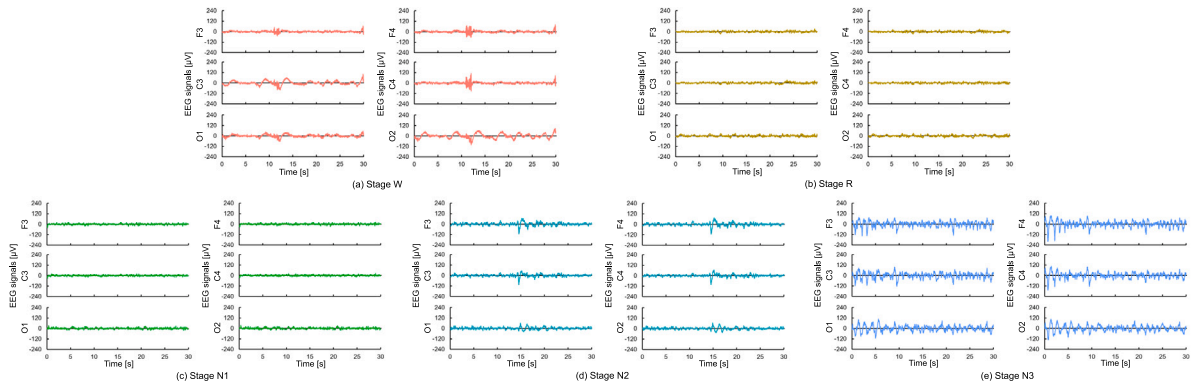

Fig. 6. Examples of measured EEG signals corresponding to each sleep stage.

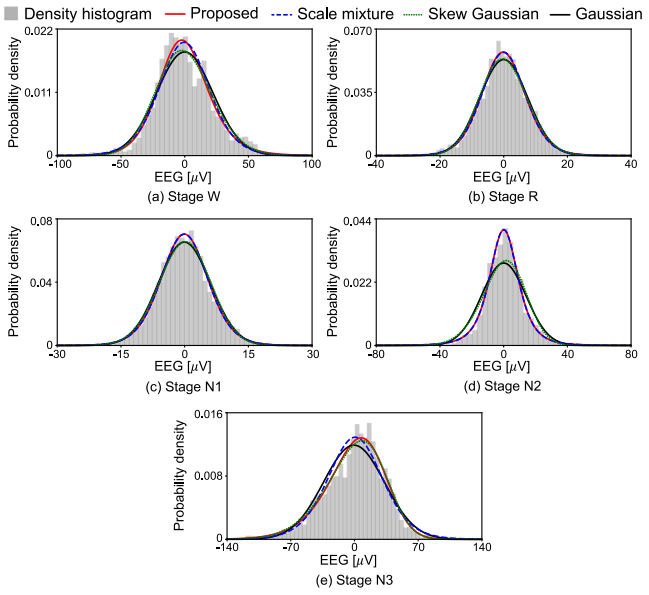

Fig. 7. Density histograms of recorded EEGs corresponding to all sleep stages. Fitted distributions estimated based on the proposed model, scale mixture model, skew Gaussian model, and Gaussian model are indicated using solid red, dashed blue, dotted green, and solid black lines, respectively.

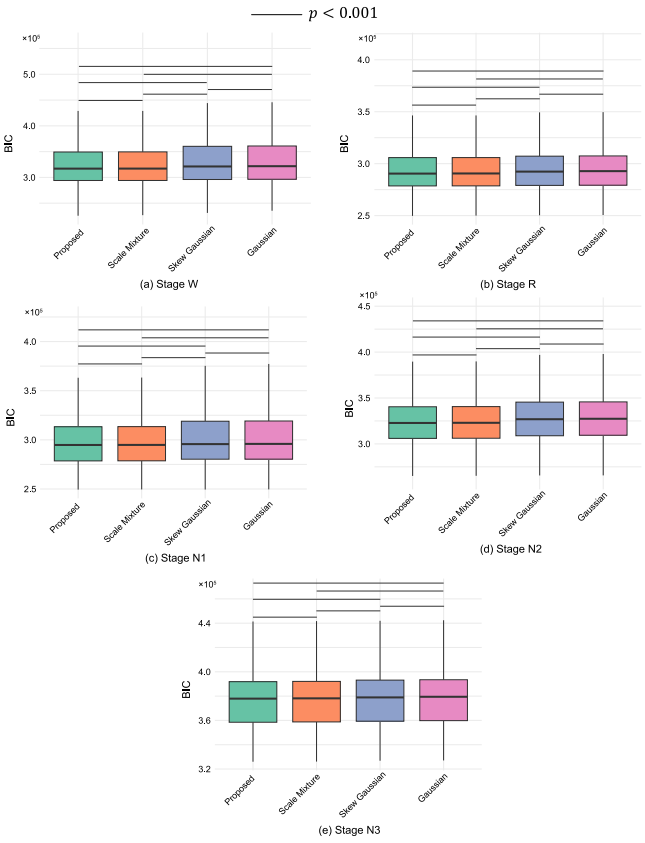

Fig. 8. Box plots showing the distribution of BIC values for Proposed model, Scale mixture model, Skew Gaussian model, and Gaussian model in each sleep stage. The statistical test results of the Wilcoxon signed-rank test are also presented.

$L = 16$ and $W = 100$ s. This suggests that calculations in high-dimensional space involve complexity, causing the computation cost to increase exponentially. The effect of changes in window length W on computation time was also evaluated. As shown in Fig. 5(e), when the window length increases, the computation time also increases, but the pattern of this increase varies depending on the number of dimensions. Specifically, when the number of dimensions L is small, the increase in computation time due to the extension of the window length is relatively gradual, whereas in high-dimensional cases, a sharp increase is observed. This indicates that the interaction between the window length and the number of dimensions significantly impacts the computation cost. Furthermore, based on the simulation results of estimation accuracy and computation time, practical scenarios for actual sleep EEG analysis were considered. For instance, assuming the use of F3, F4, C3, C4, O1 and O2 electrodes ($L = 6$) and setting the window length W to 30 s, the proposed method can estimate all parameters with an absolute percentage error of about 4% in approximately 30 s.

In the sleep EEG analysis experiment, the proposed model was found to be the most suitable model, as it consistently achieved the lowest BIC values across all sleep stages (Table 1). Furthermore, the proposed model exhibited significantly lower BIC values than conventional models, further supporting its effectiveness in sleep EEG analysis (Fig. 8). This result can be attributed to the fact that the model parameters capture both the tail weight and the lateral asymmetry of the distribution, allowing for a more flexible representation of distributional changes across sleep stage transitions. As shown in Fig. 7, EEG amplitude distributions exhibited non-Gaussian tendencies across all sleep stages, leading to high selection rates for both the proposed model and the conventional scale mixture model. Notably, in Stage N3, where the proposed model achieved the highest rate of BIC minimization, the EEG distribution was not only heavy-tailed but also asymmetric, with the tail extending more prominently in the negative direction. Since the conventional scale mixture model and the general skew Gaussian distribution can only capture either heavy tails or asymmetry independently, they were unable to fully describe such distributions. In contrast, the proposed model successfully captured

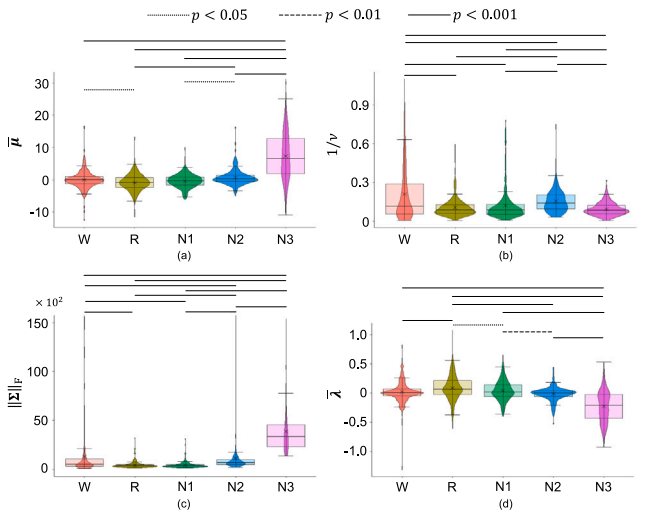

Fig. 9. Box plots and violin plots of the proposed features for all participants. (a) $\bar{\mu}$: location of the distribution. (b) $1/\nu$: non-Gaussianity corresponding to tail weight. (c) $\|\boldsymbol{\Sigma}\|_{\mathrm{F}}$: amplitude scale. (d) $\bar{\lambda}$: non-Gaussianity corresponding to lateral asymmetry. The statistical test results for pairwise comparisons of marginal means based on generalized linear models are also presented.

Table 2
Marginal means, standard errors, and 95% confidence intervals estimated using generalized linear mixed models.

Feature	Sleep stage	Estimated marginal mean	SE	Confidence interval	
				Lower	Upper
$\bar{\mu}$	W	0.017	0.26	−0.494	0.527
	R	−0.867	0.26	−1.377	−0.356
	N1	−0.369	0.26	−0.879	0.142
	N2	0.665	0.26	0.154	1.176
	N3	7.316	0.26	6.805	7.827
$1/\nu$	W	0.2001	0.01080	0.180	0.222
	R	0.1031	0.00553	0.093	0.114
	N1	0.1166	0.00628	0.105	0.130
	N2	0.1521	0.00816	0.137	0.169
	N3	0.0911	0.00489	0.082	0.101
$\|\|\boldsymbol{\Sigma}\|\|_{\mathrm{F}}$	W	1061	85.9	905	1244
	R	410	33.0	350	480
	N1	407	32.7	347	476
	N2	853	68.7	728	999
	N3	3824	307.0	3266	4476
$\bar{\lambda}$	W	0.0153	0.0129	−0.0101	0.0407
	R	0.0848	0.0129	0.0593	0.1102
	N1	0.0406	0.0129	0.0152	0.0660
	N2	−0.0191	0.0129	−0.0445	0.0063
	N3	−0.2308	0.0129	−0.2562	−0.2054

both characteristics simultaneously, making it a more comprehensive tool for sleep EEG analysis. Thus, these results demonstrate that the proposed model provides a more effective probabilistic representation of sleep EEG amplitude distributions compared to conventional models.

In Fig. 9 and Table 2, it is apparent that the proposed features depend significantly on transitions between sleep stages. The non-Gaussianity corresponding to the tail weight, $1/\nu$, was higher during Stage W and tended to increase with sleep intensity, peaking at Stage N2 and then decreasing at Stage N3. Meanwhile, $\bar{\lambda}$, which represents non-Gaussianity corresponding to the asymmetry, increased in the negative direction as sleep deepened, attaining a maximum at Stage N3. These trends are consistent with those reported in previous studies on the non-Gaussianity of sleep EEG [16,17] and are apparent from the example of the measured waveform depicted in Fig. 6 and the histogram depicted in Fig. 7.

The characteristic change in $1/\nu$ is attributed to the influence of characteristic waveforms identified during non-REM sleep. During Stage N1, a characteristic wave called the vertex sharp wave appeared around the parietal region [30]. The vertex sharp wave is a negative sharp wave that is prominently distinguished from background activity. The sudden change in amplitude associated with the appearance of this waveform may have induced a heavier tail in the distribution within the estimation window, thereby increasing $1/\nu$. Then, as sleep transitioned into Stage N2, characteristic waves called K-complex and sleep spindle appeared in the parietal region in addition to the vertex sharp waves. The K-complex is a waveform consisting of a negative sharp wave followed by a positive slow wave exceeding 200 μV, and the sleep spindle is a rhythmic wave with a frequency of 12–16 Hz [31]. The effects of these waveforms may have increased $1/\nu$ during Stage N2. Further, as sleep transitioned into Stage N3, low-frequency high-amplitude waves called δ activity became dominant. Since the percentage of δ activity increases as sleep deepens (more than 20% at Stage N3 [1]), it is possible that the relative percentage of outliers decreased in Stage N3, inducing a lighter tail of the distribution and decreasing $1/\nu$.

The change in $\bar{\lambda}$ is believed to be influenced by the hyperpolarization shift (down state) of the neuronal membrane potential during sleep. During the deep sleep stage, the duration of the down state increases, and the recorded EEG is in the form of a negative slow wave from the scalp surface [32]. Therefore, the asymmetry of the distribution is believed to have originated in Stage N2 and N3, and $\bar{\lambda}$ is believed to have increased in the negative direction.

In contrast to $\bar{\lambda}$, the location of the distribution, $\bar{\mu}$, exhibited an increasing trend with growing sleep intensity. This can be attributed to the fact that $\boldsymbol{\mu}$ compensates for the variation in the expected value of the distribution induced by the effect of $\boldsymbol{\lambda}$. From the examples depicted in Figs. 6 and 7, it is apparent that the expected value of the distribution (i.e., the mean) was approximately zero even when the distribution was deflected in the deep sleep stage. The expected value of the skew-scale mixture model is given by

$$\mathbb{E}[\mathbf{x}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\mathbb{E}\left[\frac{1}{\sqrt{u^{-1}(u^{-1}+\boldsymbol{\lambda}^{\top}\boldsymbol{\lambda})}}\right]\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\lambda}. \tag{24}$$

According to (24), the expected value of the distribution depends not only on the location parameter, $\boldsymbol{\mu}$, but also on $\boldsymbol{\lambda}$ appearing in the second term. Therefore, if the effect of the second term on the right-hand side increases because of an increase in the absolute value of $\boldsymbol{\lambda}$ induced by the skewed distribution, $\boldsymbol{\mu}$ in the first term increases in the opposite direction to ensure that the expected value of the left-hand side remains close to zero. There was no substantial change in $\|\boldsymbol{\Sigma}\|_{\mathrm{F}}$ during the stages from Stage W to N2, but it increased significantly in Stage N3. This was attributed to the high-amplitude slow-wave nature of EEG with the growing intensity of sleep [32] and the reflection of the high-amplitude feature in the distribution spread in Stage N3.

These results indicate that the proposed model can be used to evaluate changes in non-Gaussianity and amplitude scale associated with various EEG activity patterns during sleep in a unified manner. The features of non-Gaussianity proposed in this study reflect different aspects of the frequency and nonlinear dynamics of EEG that have been studied in the past. Therefore, by using the proposed features in addition to the conventional features, the proposed model may be applied to the automatic classification of sleep stages and detection of characteristic sleep waves.

7. Conclusion

We proposed a skew-scale mixture model to analyze the non-Gaussianity of sleep EEG. The proposed model can represent non-Gaussian distributions with respect to the tail weight and lateral asymmetry of the EEG distribution by introducing a skewing function

into the scale mixture model. The parameters characterizing the distribution were estimated based on marginal likelihood maximization and defined as indicators of the location, spread, and non-Gaussianity of the distribution.

Simulation experiments were conducted to verify the estimation accuracy of the proposed model parameters corresponding to varying estimation window lengths and numbers of dimensions. The results indicated that all parameters could be estimated with an error of less than 4% when the estimation window length was greater than 30 s. In the sleep EEG analysis experiment, we analyzed EEGs corresponding to Stage W, R, N1, N2 and N3. Model selection was performed for several stochastic models, suggesting that the proposed model is appropriate for EEG at all sleep stages. Further, the proposed features of non-Gaussianity, amplitude scale, and distribution location changed significantly during transitions between sleep stages. The relationship between the changes in these features and the characteristic activities of sleep EEGs suggests that the proposed model can be applied to the automatic classification of sleep stages and detection of characteristic sleep waves.

Several limitations remain in this study. First, the proposed method has limitations in its application to real-time analysis, as the computation time increases significantly depending on the window length and the number of dimensions. Particularly when using high-dimensional data or long window lengths, the computational load may become a bottleneck, potentially compromising real-time performance. Therefore, further optimization of the algorithm and the introduction of parallel computing are necessary.

Second, while this study conducted simulation experiments and analyzed sleep EEG data to verify the accuracy and effectiveness of the proposed method, it did not specifically evaluate its application to practical tasks such as the automatic classification of sleep stages or the detection of characteristic sleep waves. Further validation and improvement in these practical applications are required in future research.

CRediT authorship contribution statement

Miyari Hatamoto: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Akira Furui:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Keiko Ogawa:** Writing – review & editing, Resources, Investigation, Data curation. **Toshio Tsuji:** Writing – review & editing, Supervision, Project administration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Akira Furui reports financial support was provided by Japan Society for the Promotion of Science. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Equivalent expression for joint distribution

This appendix shows that the joint distributions given by (10) and (11) are equivalent. From (10), the joint distribution of $\mathbf{x}_n, u_n, \tau_n$ is given by

$$
\begin{aligned}
&p(\mathbf{x}_n, u_n, \tau_n) \\
&= \mathcal{N}\left(\mathbf{x}_n|\boldsymbol{\mu} + \tau_n\sqrt{u_n}\boldsymbol{\Sigma}^{\frac{1}{2}}\delta_{u_n}, u_n\boldsymbol{\Sigma}^{\frac{1}{2}}\left(\mathbf{I}_L + \lambda_{u_n}\lambda_{u_n}^{\mathsf{T}}\right)^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\right) \\
&\quad \times \mathcal{H}\mathcal{N}(\tau_n|0,1)\mathrm{IG}(u_n|\nu/2, \nu/2). \qquad (25)
\end{aligned}
$$

Here, the probability density function of the half-Gaussian distribution can be expressed as $\mathcal{H}\mathcal{N}(\tau_n|0,1) = 2\mathcal{N}(\tau_n|0,1)$ for $\tau_n \in [0,\infty)$. Further, by substituting $\mathbf{A} = u_n\boldsymbol{\Sigma}^{\frac{1}{2}}\delta_{u_n}$, $\boldsymbol{\Sigma}_a = u_n\boldsymbol{\Sigma}^{\frac{1}{2}}(\mathbf{I}_L + \lambda_{u_n}\lambda_{u_n}^{\mathsf{T}})^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}$ and $\Lambda = (1 + \mathbf{A}\boldsymbol{\Sigma}_a^{-1}\mathbf{A}^{\mathsf{T}})$, (25) can be transformed as follows [21]:

$$
\begin{aligned}
&p(\mathbf{x}_n, u_n, \tau_n) \\
&= 2\mathcal{N}\left(\mathbf{x}_n|\boldsymbol{\mu} + \tau_n\mathbf{A}, \boldsymbol{\Sigma}_a\right)\mathcal{N}(\tau_n|0,1)\mathrm{IG}(u_n|\nu/2,\nu/2) \\
&= 2\mathcal{N}\left(\mathbf{x}_n|\boldsymbol{\mu}, \boldsymbol{\Sigma}_a + \mathbf{A}\mathbf{A}^{\mathsf{T}}\right)\mathcal{N}(\tau_n|\Lambda\mathbf{A}^{\mathsf{T}}\boldsymbol{\Sigma}_a^{-1}(\mathbf{x}_n - \boldsymbol{\mu}), \Lambda) \\
&\quad \times \mathrm{IG}(u_n|\nu/2,\nu/2) \\
&= 2\mathcal{N}\left(\mathbf{x}_n|\boldsymbol{\mu}, u\boldsymbol{\Sigma}\right)\mathcal{N}(\tau_n|\Lambda^{\frac{1}{2}}\lambda^{\mathsf{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}_n - \boldsymbol{\mu}), \Lambda) \\
&\quad \times \mathrm{IG}(u_n|\nu/2,\nu/2) \\
&= 2\mathcal{N}\left(\mathbf{x}_n|\boldsymbol{\mu}, u\boldsymbol{\Sigma}\right)\mathcal{N}(\tau_n|\lambda^{\mathsf{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x}_n - \boldsymbol{\mu}), 1) \\
&\quad \times \mathrm{IG}(u_n|\nu/2,\nu/2). \qquad (26)
\end{aligned}
$$

This proves that (25) and (11) are equivalent.

Data availability

The authors do not have permission to share data.

Skew Scale Mixture Model (Original data) (Mendeley Data)

References

[1] C. Iber, S. Ancoli-Isreal, A. Chesson, S. Quan, The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Techinical Specifications, Westchester II, 2007.

[2] J.G. Klinzing, N. Niethard, J. Born, Mechanisms of systems memory consolidation during sleep, Nat. Neurosci. 22 (10) (2019) 1598–1610.

[3] S. Chouinard, J. Poulin, E. Stip, R. Godbout, Sleep in untreated patients with schizophrenia: a meta-analysis, Schizophr. Bull. 30 (4) (2004) 957–967.

[4] Y. Wei, M.A. Colombo, J.R. Ramautar, T.F. Blanken, Y.D. van der Werf, K. Spiegelhalder, B. Feige, D. Riemann, E.J.W. Van Someren, Sleep stage transition dynamics reveal specific stage 2 vulnerability in Insomnia, Sleep 40 (9) (2017).

[5] K. Šušmáková, Human sleep and sleep EEG, Meas. Sci. Rev. 4 (2) (2004) 59–74.

[6] C. Stepnowsky, D. Levendowski, D. Popovic, I. Ayappa, D.M. Rapoport, Scoring accuracy of automated sleep staging from a bipolar electroocular recording compared to manual scoring by multiple raters, Sleep Med. 14 (11) (2013) 1199–1207.

[7] W. Al-Salman, Y. Li, P. Wen, Detecting sleep spindles in EEGs using wavelet fourier analysis and statistical features, Biomed. Signal Process. Control. 48 (2019) 80–92.

[8] Y. You, X. Zhong, G. Liu, Z. Yang, Automatic sleep stage classification: A light and efficient deep neural network model based on time, frequency and fractional Fourier transform domain features, Artif. Intell. Med. 127 (102279) (2022) 102279.

[9] M. Shahbakhti, M. Beiramvand, T. Eigirdas, J. Sole-Casals, M. Wierzchon, A. Broniec-Wojcik, P. Augustyniak, V. Marozas, Discrimination of wakefulness from sleep stage I using nonlinear features of a single frontal EEG channel, IEEE Sens. J. 22 (7) (2022) 6975–6984.

[10] S. Tiwari, D. Arora, V. Nagar, Supervised approach based sleep disorder detection using non - Linear dynamic features (NLDF) of EEG, Meas. Sens. 24 (100469) (2022) 100469.

[11] P. An, J. Zhao, B. Du, W. Zhao, T. Zhang, Z. Yuan, Amplitude-time dual-view fused EEG temporal feature learning for automatic sleep staging, IEEE Trans. Neural Netw. Learn. Syst. 35 (5) (2024) 6492–6506.

[12] P. Mamta, S.V.A.V. Prasad, EEG feature extraction using time domain analysis for classifying insomnia, in: 2020 IEEE India Council International Subsections Conference, INDISCON, IEEE, 2020, pp. 221–224.

[13] A. Furui, R. Onishi, A. Takeuchi, T. Akiyama, T. Tsuji, Non-Gaussianity detection of EEG signals based on a multivariate scale mixture model for diagnosis of epileptic seizures, IEEE Trans. Biomed. Eng. 68 (2) (2021) 515–525.

[14] A. Furui, R. Onishi, T. Akiyama, T. Tsuji, Epileptic seizure detection using a recurrent neural network with temporal features derived from a scale mixture EEG model, IEEE Access 12 (2024) 162814–162824.

[15] S. Fukuda, A. Furui, M.G. Machizawa, T. Tsuji, Stochastic fluctuation in EEG evaluated via scale mixture model for decoding emotional valence, in: 2024 IEEE/SICE International Symposium on System Integration, SII, IEEE, 2024, pp. 567–572.

[16] M.S. Weiss, Non-Gaussian properties of the EEG during sleep, Electroencephalogr. Clin. Neurophysiol. 34 (2) (1973) 200–202.

[17] H. Sugimoto, N. Ishii, A. Iwata, N. Suzumura, T. Tomita, On the stationarity and normality of the electroencephalographic data during sleep stages, Comput. Programs Biomed. 8 (3–4) (1978) 224–234.
[18] S. Gune, K. Polat, M. Dursun, S. Yosunkaya, Examining the relevance with sleep stages of time domain features of EEG, EOG, and chin EMG signals, in: Proc. of 14th BIYOMUT, 2009, pp. 1–4.
[19] C. Metzner, A. Schilling, M. Traxdorf, H. Schulze, P. Krauss, Sleep as a random walk: a super-statistical analysis of EEG data across sleep stages, Commun. Biol. 4 (1) (2021) 1385.
[20] M. Hatamoto, A. Furui, K. Ogawa, T. Tsuji, Sleep EEG analysis based on a scale mixture model and its application to sleep spindle detection, in: 2022 IEEE/SICE International Symposium on System Integration, SII, ieeexplore.ieee.org, 2022, pp. 887–892.
[21] C.S. Ferreira, V.H. Lachos, H. Bolfarine, Likelihood-based inference for multivariate skew scale mixtures of normal distributions, AStA Adv. Stat. Anal. 100 (4) (2016) 421–441.
[22] C.M. Bishop, Pattern Recognition and Machine Learning, springer, 2006.
[23] J.A. Bilmes, et al., A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, Int. Comput. Sci. Inst. 4 (510) (1998) 126.
[24] G.H. Golub, C.F. Van Loan, Matrix Computations, JHU Press, 2012.
[25] C. O'Reilly, N. Gosselin, J. Carrier, T. Nielsen, Montreal Archive of Sleep Studies: an open-access resource for instrument benchmarking and exploratory research, J. Sleep Res. 23 (6) (2014) 628–635.
[26] G. Schwarz, Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[27] A. Azzalini, A.D. Valle, The multivariate skew-normal distribution, Biometrika 83 (4) (1996) 715–726.
[28] D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear Mixed-Effects models using lme4, 2014, arXiv:1406.5823.
[29] R. Lenth, H. Singmann, J. Love, P. Buerkner, M. Herve, Emmeans: Estimated marginal means, aka least-squares means, 2018, R Package Version 1. 3. 0.
[30] T. Hori, M. Hayashi, T. Morikawa, Topographical EEG changes and the hypnagogic experience, in: R.D. Ogilvie (Ed.), Sleep Onset: Normal and Abnormal Processes, Vol. 397, American Psychological Association, xxviii, Washington, DC, US, 1994, pp. 237–253.
[31] T. Hori, Y. Sugita, E. Koga, S. Shirakawa, K. Inoue, S. Uchida, H. Kuwahara, M. Kousaka, T. Kobayashi, et al., Proposed supplements and amendments to 'a manual of standardized terminology, techniques and scoring system for sleep stages of human subjects', the Rechtschaffen & Kales (1968) standard, Psychiatry Clin. Neurosci. 55 (3) (2001) 305–310.
[32] G. Buzsaki, Rhythms of the Brain, Oxford University Press, 2006.