A data assimilation method to track time-varying changes in the excitation-inhibition balance using scalp EEG

Hiroshi Yokoyama[1,2,3*] and Keiichi Kitajo[1,2*]

[1*]Division of Neural Dynamics, National Institute for Physiological Sciences, 38 Nishigonaka, Myodaiji, Okazaki, Aichi, 444-8585, Japan.

[2]Department of Physiological Sciences, School of Life Science, Graduate University for Advanced Studies (SOKENDAI), 38 Nishigonaka, Myodaiji, Okazaki, Aichi, 444-8585, Japan.

[3] Program of Mathematical and Life Sciences, Graduate School of Integrated Sciences for Life, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima, 739-8526, Japan.

*Corresponding author(s). E-mail(s): yokoyama@nips.ac.jp; hyokoyama@hiroshima-u.ac.jp; kkitajo@nips.ac.jp;

Abstract

Recent neuroscience studies have suggested that control of the excitation and inhibition (E/I) balance is important to maintain normal brain function. However, an efficient method to evaluate the time-varying changes in E/I balance of the brain has yet to be established. To tackle this issue, we propose a new approach to estimate E/I balance changes, by applying the method of neural-mass model-based tracking of the brain state using the Ensemble Kalman Filter. In this method, the parameters regarding the synaptic E/I gains of the model are estimated from observed electroencephalography (EEG). Moreover, the index of E/I balance was defined by calculating the ratio between synaptic E/I gains based on estimated parameters. To validate this method, we applied it to numerical and

human EEG data. As a result, we confirmed that our proposed method could estimate the E/I balance changes from observed human EEG.

Keywords: Data assimilation, E/I balance, Neural-mass model, electroencephalography, Ensemble Kalman Filter

Introduction

A balance between synaptic excitation and inhibition (E/I) plays a crucial role in the neural mechanisms underlying social behavior, neuropsychiatric disorder-related symptoms, cognitive performance, sleep function, and synaptic homeostasis. For example, some animal studies have shown that synaptic E/I imbalance within the prefrontal cortex leads to autistic-like social dysfunction [1]. In the context of human neuroscience, studies using concurrent transcranial magnetic stimulation and electroencephalography (TMS-EEG) recordings have reported evidence of a relationship between cortical GABAergic inhibition/glutamatergic excitation and psychiatric disorders [2, 3]. Therefore, the disruption of E/I balance in the human brain could also lead to social dysfunction. Furthermore, both animal and human studies have suggested that changes in E/I balance within the visual cortex underly the regulation of sleep-dependent changes in synaptic excitation [4, 5]. Establishing a data-driven method to detect temporal changes in E/I balance within the intact human brain would allow us to elucidate the functional role of dynamics in synaptic E/I in the brain. While techniques to quantify the E/I balance in the human brain in vivo have been developed (e.g., MRS: magnetic resonance spectroscopy [5, 6]), these methods cannot evaluate temporal changes in E/I balance at a sub-minute scale.

The changes in E/I balance could be considered as the changes in synaptic current mediated by the glutamatergic excitation and GABAergic inhibition.

Thus, we expect that the oscillatory activity in human scalp EEG reflects the temporal dynamics of E/I balance. In fact, recent TMS-EEG studies suggested that TMS-evoked potential (TEP) in scalp EEG reflect both GABAergic and glutamatergic mediated functions [2, 3, 7, 8]. However, to our knowledge, no study has attempted to directly quantify the neural features of this synaptic E/I using only observed EEG signals. To this aim, we propose a new model-based method to estimate temporal changes in E/I balance from observed EEG signals by applying neural-mass (NM) model-based tracking [9, 10] of the brain state using the Ensemble Kalman Filter (EnKF) [11, 12]. The EnKF was developed to assimilate the nonlinear dynamical models with observed data. In the field of control theory, the nonlinear Kalman Filter scheme (including the EnKF) is applied not only to predict time-series data but also to estimate model parameters. Conversely, mathematical studies in neuroscience have reported that the time-evolving dynamics in EEG signals can be formulated as a nonlinear dynamical system, such as that predicted by the NM model [9]. Moreover, the NM model contains parameters of the average synaptic gain for excitatory and inhibitory interneurons. To summarize, we predicted that temporal changes in glutamatergic excitation and GABAergic inhibition can be estimated from observed EEG via parameter estimation of the NM model using the EnKF scheme. In addition, we propose an evaluation index of E/I balance changes, which is determined by calculating the ratio between synaptic E/I gain parameters in the NM model.

To validate our proposed method, two datasets were analyzed. First, we applied our proposed method to the synthetic data generated by the NM model with known temporal changes in the model parameters. By doing so, we confirmed whether the estimated model parameters by using our method are consistent with exact model parameters. Next, we applied our method to

an open dataset of experimental EEG signals measured from healthy humans while sleeping [13]. In this way, we tested whether our proposed method could detect the time-varying changes in E/I balance during sleep from observed EEG signals.

Results

An overview of the proposed method

In the proposed method, single-channel time-series data of observed EEG signals are sequentially fitted with an NM model [9] using variational Bayesian noise adaptive constrained EnKF (vbcEnKF) on a sample-by-sample basis (see the Methods section for details). Moreover, five model parameters, A, a, B, b, and p, in the NM model (see the Methods section for parameter descriptions) were also reconstructed in parallel with predicting the changes in EEG signals in the vbcEnKF scheme. Here, we assumed that changes in the E/I balance are reflected in the changes in model parameters regarding the E/I synaptic gain parameters A and B. This being the case, the neural feature of E/I balance can be quantified by calculating the ratio of these parameters. Therefore, we proposed the model-based evaluation index of E/I ratio (mE/I ratio: $\mathrm{mE/I} = A/(A+B)$). By applying the proposed method, we can directly estimate the time-varying changes in E/I ratio from observed EEG signals (see the Methods section and Supplementary information for more details). A schematic image of this proposed method is shown in Fig. 1.

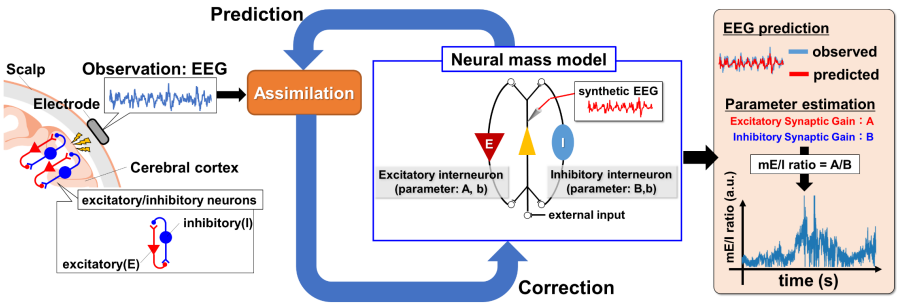


Fig. 1 Overview of the proposed method. The single-channel time-series data of observed EEG were sequentially predicted with a neural-mass (NM) model using the variational Bayesian noise adaptive constrained Ensemble Kalman Filter (vbcEnKF), and five model parameters, A, a, B, b,and p, were also estimated. Based on our assumption that the changes in excitation-inhibition (E/I) balance are reflected in changes of the model parameters A and B, we proposed the model-based evaluation index of the E/I ratio (mE/I ratio= $A/(A+B)$).

Verification with numerical simulation

To confirm whether our proposed method can correctly estimate the temporal changes in model parameters from the observed data, we applied our method to synthetic data generated by an NM model with known parameter changes. Moreover, as described in the Methods section, our proposed method applied the combination approach of EnKF with a variational Bayesian noise adaptive algorithm [14–18] to avoid reduction in estimation accuracy caused by non-stationarity of the observation noise. Therefore, we also evaluated whether the estimated covariance of observation noise was consistent with exact noise covariance containing the synthetic EEG signal. In addition, since the EnKF is based on the sequential Monte Carlo (sMC) method for Bayesian probability estimation, both the initial seed value and ensemble size for the random sampling should affect the error of state estimation. Therefore, to compare the prediction error according to the ensemble size and initial random seed of sMC sampling, we selected ensemble sizes from 40 to 500 with 20 steps and applied these to the same synthetic data 50 times

with a different initial random seed for each ensemble size (see the Methods section for a more detailed description of simulation settings).

The results of this validation are shown in Figs. 2 and 3. Fig. 2 shows examples of state estimation results with $N_{ens} = 40$ and 200. While the predicted EEG time series seemed to be qualitatively similar to the exact synthetic EEG between both ensemble size conditions (Fig. 2a, d), the estimation of the mE/I ratio tended to be accurate in relation to the increase of ensemble size (Fig. 2b, e). Moreover, the estimation of the model parameters also tended to be accurate depending on the increase of ensemble size (Fig. 2c, f). To quantitatively reveal the prediction skill according to the ensemble size in our method, we then assessed the accuracy of EEG signal prediction and observed the estimated noise covariance (Fig. 3). As shown in Fig. 3a, the prediction error of the EEG signal, based on the mean absolute error score, showed a tendency toward decreasing the error with an increase in the ensemble size N_{ens}. This tendency was also found in the estimation results of observation noise covariance (Fig. 3b). Moreover, estimated noise covariance converged to exact noise covariance with $N_{ens} \geq 200$.

In summary, the proposed approach was able to predict the model states and estimate parameters, and $N_{ens} \geq 200$ was required to guarantee an accurate prediction with a smaller prediction error.

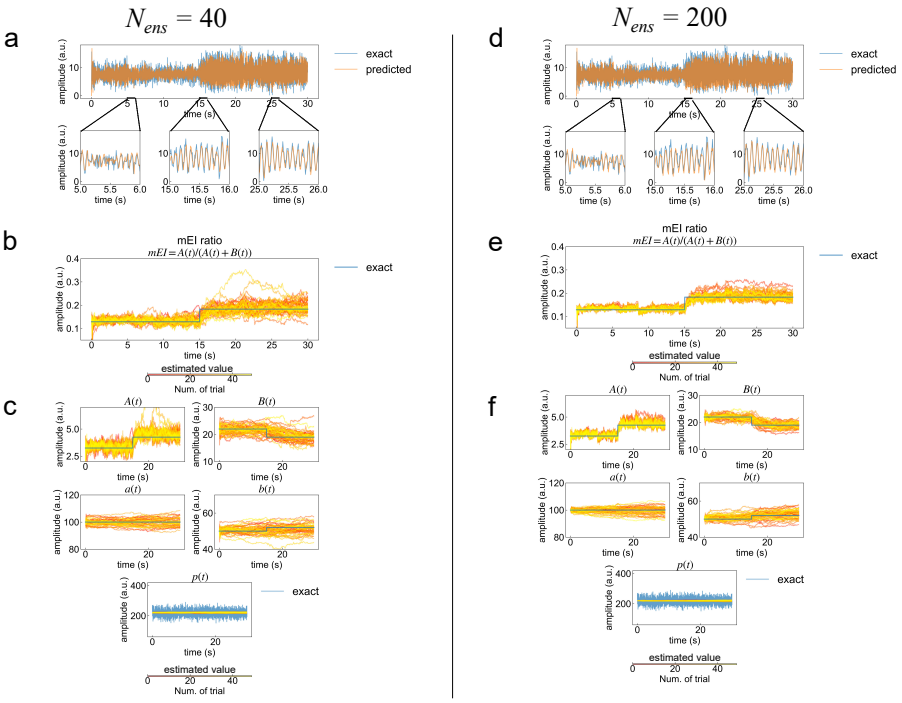


Fig. 2 Time-series prediction of EEG data and estimation of unobservable model parameters using synthetic EEG data. a, d) Comparison of original EEG and typical prediction result of EEG obtained from 50 trials with an ensemble size of $N_{ens} = 40$ and 200, respectively. The upper panels in (a) and (d) show the whole time series of synthetic and predicted EEG. The lower three panels in (a) and (d) are enlarged views of the prediction results for each time interval. Blue and orange lines indicate the time series of exact and predicted values, respectively. b, e) Estimation results of the mE/I ratio obtained from 50 trials with an ensemble size of $N_{ens} = 40$ and 200, respectively. c, f) Estimation results of the five parameters, A, a, B, b, and p, obtained from 50 trials with an ensemble size of $N_{ens} = 40$ and 200, respectively. Blue lines in (b)–(f) indicate the time series of exact values. Orange gradient color lines in (b)–(f) indicate the estimated value of each parameter obtained from 50 trials.

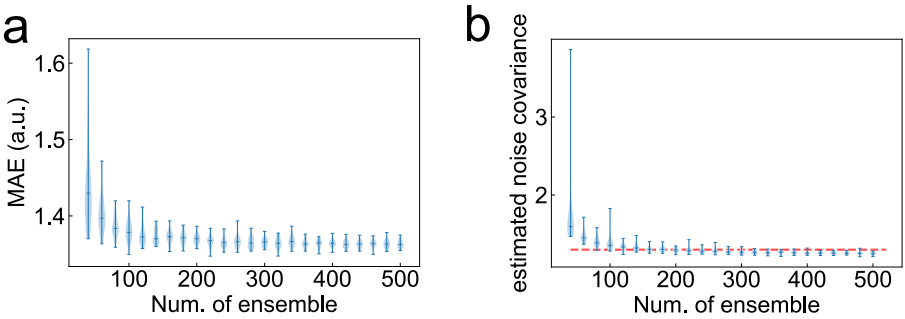


Fig. 3 Effects of ensemble size on estimation accuracy. a) Violin plots of prediction error scores as a function of the number of ensembles. The probability densities in these violin plots were estimated using a kernel density estimation method based on the samples obtained from 50 trials. The error bar indicates the maximum and minimum value of samples. The middle line of the error bar shows the median value of the samples. b) Corresponding estimated observation noise covariance as a function of the number of ensembles. Red-dot line indicates the exact noise covariance of synthetic EEG data.

Sleep EEG analysis

To confirm the neurophysiological validity of our proposed method, we applied it to real human EEG data obtained from Kemp et al. (2000) [13]. This dataset includes EEG recorded during day-night periods (over around 20 hours) from healthy participants at their home. The sleep-stage markers were scored manually by referring to recorded signals, and a more detailed description of the task settings is provided in the original paper [13]. In general, when sleeping, EEG oscillations between the delta and lower beta frequency band change depending on the sleep stages, such as nonrapid eye movement (NREM) or rapid eye movement (REM) sleep. Moreover, recent animal and human studies in sleep neurophysiology have suggested that sleep regulates E/I balance [4, 5]. In particular, Tamaki et al. (2020) [5] reported that the changes in the E/I balance in the early visual area are sleep-stage dependent. Using simultaneous magnetic resonance spectroscopy (MRS), the authors found that modulation of the E/I balance played a role in stabilizing visual learning. Therefore, we applied our method to sleep EEG data to

determine whether it could detect changes in E/I balance depending on the NREM/REM sleep stage from observed human EEG signals. For this analysis, we used EEG data from 19 healthy participants who showed typical transitions between the sleep stages during the first NREM (Stage 1, Stage 2, and Stage 3/4) and REM periods. Moreover, in the following validation tests, we used an EEG interval from the first 20-minute period before sleep onset to the end of the first REM period for each participant. In the following analysis, the interval from the first 20-minute period before the sleep onset was defined as the awake period.

A typical result of the time-series prediction of EEG in one participant (ID: 4032) is shown in Fig. 4. Our method correctly tracked the time-evolving dynamics of the original EEG signal within the sleep and awake intervals (Fig. 4 b–d). The estimation results of time-varying changes in the mE/I ratio for two typical participants are shown in Fig. 5. While the estimated E/I ratio in the NREM period increased (i.e., synaptic gain of excitatory interneurons tended to be dominant during NREM periods), the value in the REM period decreased (i.e., synaptic gain of excitatory interneuron tended to be suppressed during REM periods). This result is consistent with the experimental results of an MRS study conducted by Tamaki et al. (2020) [5]. Moreover, the estimated E/I ratio in the awake period was also smaller than that in the NREM period, and this was seen in both participants' data (see Fig. 5a, b). A similar tendency has been reported in an animal study [4].

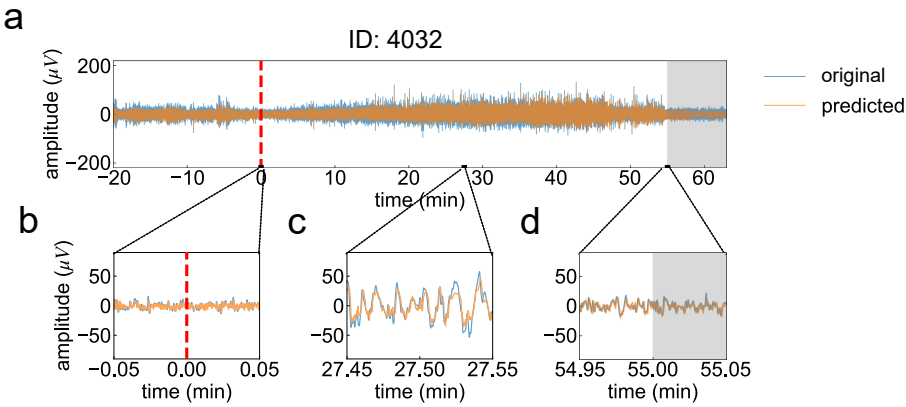


Fig. 4 Typical examples of predicted EEG signals. a) Prediction result of an EEG signal. Blue and orange lines indicate the time series of observed and predicted EEG signals, respectively. On the x-axes, the position at time = 0 (min) indicates the sleep onset (onset of NREM period). The gray shaded area indicates the REM sleep period. b–d) Enlarged views of forecasting signals with a 0.1 min time window for three-time intervals around sleep onset (b), middle of NREM (c), and around the end of NREM (i.e., the onset of REM) periods (d). The sleep stages in this dataset were manually determined by well-trained technicians [13].

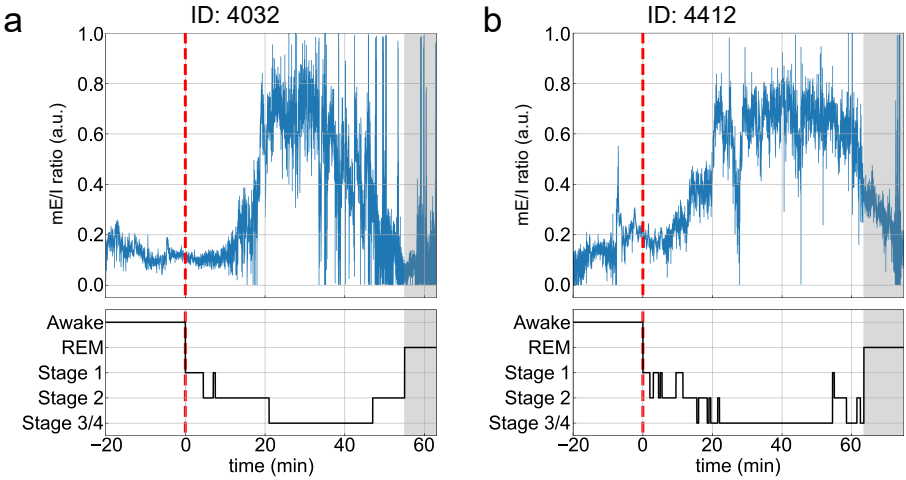


Fig. 5 Typical examples of time-varying changes in the mE/I ratio. a, b) Results for two participants. The upper panels of (a) and (b) show the time series of the mE/I ratio estimated by our proposed method from the observed EEG signals. Blue lines indicate the time series of the mE/I ratio. The lower panels in (a) and (b) show the time profiles of sleep-stage changes. Red dotted lines indicate the sleep onset (onset of NREM period). Gray shaded areas indicate the REM period.

Next, to address the statistical validity of the E/I results, the time-averaged mE/I ratio of each sleep stage (Awake, State 1, State 2, State 3/4, and NREM) were evaluated for all 19 participants (Fig. 6a). Then, the group data of these time-averaged mE/I ratios were subjected to a Kruskal-Wallis test, which is a rank-based nonparametric version of a one-way ANOVA. The Kruskal-Wallis test revealed that the time-averaged mE/I ratios were significantly different across distinct sleep stages (h-statistics $H = 32.1325$, $p < 0.0001$, effect size $\eta^2 = 0.3126$). Furthermore, by subjecting these time-averaged mE/I ratios to the Dunn's multiple comparison test (two-tailed test), which is a rank-based method for post-hoc comparisons, the time-averaged mE/I ratio in Stage 3/4 (i.e., NREM Stage 3/4) was significantly higher than those in the Awake, Stage 2, and REM periods (Stage 3/4 vs. Awake, $p < 0.0001$; Stage 3/4 vs. Stage 2, $p = 0.001$; Stage 3/4 vs. REM, $p = 0.0057$; p-values were corrected using the Bonferroni method; Fig. 6b). Moreover, the mE/I ratio in Stage 2 (i.e., NREM Stage 2) was significantly higher than in the awake period ($p = 0.007$). Other paired comparisons of sleep stages showed no significant differences for the time-averaged mE/I ratios (Awake vs. Stage 1, $p = 1.00$; Awake vs REM, $p = 1.00$; Stage 1 vs. Stage 2, $p = 0.16$; Stage 1 vs. REM, $p = 1.00$; Stage 2 vs. Stage 3/4, $p = 1.00$; Stage 2 vs. REM, $p = 0.52$).

In summary, these results indicated that changes of the mE/I ratio estimated by our proposed method using observed EEG signals were consistent with the experimental evidence reported by prior studies [4, 5].

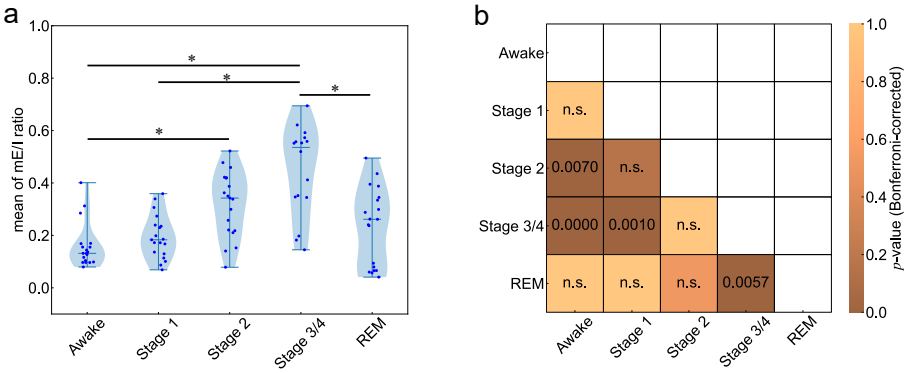


Fig. 6 Sleep-state dependent changes in the mE/I ratio. a) The violin plots were obtained from the samples of the mE/I ratio in 19 participants for each sleep stage. The error bar indicates the maximum and minimum value of samples. The middle line of the error bar shows the median value of the samples. The blue dots indicate the samples of the time-averaged mE/I ratios for each participant. Asterisks (*) indicate $p < 0.01$ by the Dunn's multiple comparison test with Bonferroni correction. b) Multiple comparisons of each pair of mean mE/I ratios by the sleep stage. p-values were adjusted using the Bonferroni method.

Discussion

In this study, we proposed a new approach to track the time-varying changes in E/I balance from observed EEG data on a sample-by-sample basis. Using both numerical and empirical neurophysiological data, our validation results supported our assumptions, as follows: (1) Temporal changes in the E/I ratio caused by synaptic current changes are reflected in the dynamics of scalp EEG, and (2) temporal changes in the E/I ratio could be estimated using estimated parameters in the NM model with the EnKF scheme. Moreover, sleep stage-dependent changes in the mE/I ratio estimated from sleep EEG data (see Figs. 5 and 6) were consistent with previous experimental MRS results reported by Tamaki et al. (2020) [5]. Based on these previous results and our own, we suggest that the proposed method could be applied to human EEG to reveal the sleep-related mechanisms underlying changes in the E/I ratio. Moreover, while MRS cannot track time-varying changes in the E/I ratio because of its inherent measurement limitations, our method can estimate such time-varying changes using the proposed metric, i.e., the mE/I ratio. This difference in the temporal resolution for tracking E/I balance changes is the most important advantage of our proposed method.

In the context of human neuroscience, some experimental evidence based on TMS-EEG techniques has suggested that TEP changes recorded on human scalp EEGs reflect both GABAergic-mediated inhibitory and glutamatergic-mediated excitatory functions [7, 8]. Therefore, by evaluating the TEP based on TME-EEG techniques, we could determine how intracortical E/I reflects neural mechanisms underlying the neuropsychiatric disorder associated with E/I balance changes [2, 3]. TEP is assessed by calculating the averaged EEG signals measured during experimental tasks along with TMS stimulation over multiple trials, and TMS-EEG-based

evaluations for intracortical E/I require several trials to define the TEP component of EEG signals. Therefore, although TEP is a cutting-edge measure, it cannot conveniently reveal time-varying changes in E/I balance. In contrast to this disadvantage of the TMS-EEG-based method, the current results suggest that our proposed method can overcome this issue, such as by directly estimating time-varying changes in the E/I ratio from observed EEG data on a single-trial basis. The originality of our proposed method lies in that the method can reconstruct unobserved features of the E/I ratio from observed EEG on a single-trial basis by applying data assimilation techniques.

In another context of methodological study on local field potentials (LFPs), one study proposed a time-frequency analysis-based approach to track temporal changes in E/I balance from LFP signals [19]. This study demonstrated that the slope of the log-log plot of the power spectral density functions between 30 and 50 Hz in LFPs (i.e., $1/f$ power law exponent in the gamma frequency band) was correlated with the E/I ratio. Therefore, the authors proposed a power spectral density-based E/I estimation method (hereinafter called the E/I slope method). Moreover, a recent study [20] applied E/I slope analysis to human EEG data and found a significant difference in the E/I balance between awake and REM states. To compare with our proposed method, we applied this E/I slope method to the same sleep EEG datasets [13] used in the present study. The results are shown in the Supplementary information (Supplementary Figs. 1 and 2). However, the sleep-dependent changes in the E/I ratio were not observed when using the E/I slope method, unlike previous findings [20]. No significant difference in the E/I slope depending on the sleep stages was found (Supplementary Fig. 2). The failure of the E/I slope method to detect sleep-stage changes in our dataset may be because gamma activity (> 30 Hz) in EEG tends to be

compromised due to electromyogenic artifacts from cranial and ocular muscles. Moreover, since the sampling frequency of EEG signals in the dataset we used was 100 Hz (i.e., the meaningful frequency components of observed EEG in our dataset were below the 50 Hz), the signal of gamma band oscillations over 50 Hz would not be properly observed in this dataset. This could be why the E/I slope method cannot efficiently detect the sleep-dependent changes in E/I balance from EEG data. In practice, power spectra in the gamma band (especially those greater than 50 Hz) of our EEG data were weaker than those of lower frequency bands. In contrast, our proposed method can detect significant differences in the mE/I ratio between sleep stages, even for situations in which the E/I slope method cannot (Figs. 5 and 6). Our method directly focused on frequency bands that indicate sleep-dependent oscillations under 20 Hz (i.e., delta, theta, and alpha oscillations) to evaluate the mE/I ratio from observed EEG data with the vbcEnKF scheme in the NM model. This contrast between our proposed and prior methods highlights the advantage of our method, in that it can estimate the E/I ratio from observed EEG data without relying on signal quality or sampling frequency. Moreover, by directly assimilating the observed EEG to the NM model, we can interpret how sleep-dependent EEG oscillations are caused by the synaptic balance between excitation and inhibition according to changes in the estimated E/I ratio. This is another advantage of our proposed method.

To the best of our knowledge, the first study to adapt data assimilation techniques in neuroscience research was conducted by Ullah et al. (2010) [21]. That study reported that an unobserved process of neuronal dynamics associated with cellular excitability during seizure can be reconstructed from a single observed membrane potential by assimilating the neuron model behavior to real observed data. In the past decade, some other studies have

also used data assimilation techniques [10, 21–23]. In particular, Kuhlmann et al. (2016) [10] proposed that the depth of anesthesia can be tracked by looking at the changes in predicted model parameters of the NM model from observed human EEG using the unscented Kalman Filter. These prior studies applied the state and parameter estimation approach using the nonlinear Kalman Filter or Bayesian filter for the data assimilation of neural data, as with our proposed method. However, those previous methods did not consider the effect of nonstationary observation noise on the accuracy of model parameter estimation in the Kalman Filter scheme. In addition, most neurophysiological models contain the parameter constraints to allow for stable dynamics of its model behavior; however, the prior methods did not consider this issue when assimilating the model prediction to observed data. Comparing our proposed method with those of the above-mentioned studies, through use of the vbcEnKF scheme, our method can track the time-evolving dynamics of observed signals while considering both the effect of nonstationary observation noise and parameter constraints in the model. As shown in the results of numerical simulations, since the vbcEnKF can correctly determine the observation noise scale with the ensemble size $N_{ens} \geq 200$, our proposed method can adaptively predict the EEG time series and E/I ratio while overcoming the effects of nonstationary observation noise. These significant advantages of our proposed method are not seen in those previous conventional methods.

Despite the advantages of our proposed method, some limitations exist. First, our proposed method can be applied only for single-channel EEG signal, that is, the proposed method cannot be applied to estimate the network-level dynamics of EEG signals. Even in the case of the parameter estimation for a single-column NM model from observed EEG, we must determine 11 state

variables (six state variables and five model parameters) using the EnKF. Therefore, if we are to estimate the network-level dynamics of E/I balance changes from multiple-channel EEG data, an efficient parameter estimation algorithm with the EnKF scheme for high-dimensional data needs to be developed. Further studies are required to address this issue. Second, our proposed method remains the effect of the initial setting for state noise covariance Q. To avoid reducing estimation performance with nonstationary observed noise, the scale of observation noise covariance is recursively optimized using the variational Bayesian noise adaptation method. However, the scale of state noise covariance Q is fixed in our proposed method. To consider the effects of nonstationarity in both state and observed noise, the variational Bayesian algorithm we applied in this method should be modified.

Methods

Definition of the NM model

The NM model, proposed by Jansen and Rit [9], is a mathematical model for EEG that is formulated with the evolution of the average postsynaptic potential (PSP) in the three following interacting neural populations: pyramidal cells, excitatory interneurons, and inhibitory interneurons. This model can be described using the following six first-order differential

equations [9]:

$$
\begin{aligned}
\dot{v}_0(t) &= v_3(t) \\
\dot{v}_1(t) &= v_4(t) \\
\dot{v}_2(t) &= v_5(t) \\
\dot{v}_3(t) &= Aa\mathrm{Sigm}(v_1(t) - v_2(t)) - 2av_3(t) - a^2 v_0(t) \\
\dot{v}_4(t) &= Aa[p(t) + C_2\mathrm{Sigm}(C_1 v_0(t))] - 2av_4(t) - a^2 v_1(t) \\
\dot{v}_5(t) &= BbC_4\mathrm{Sigm}(C_3 v_0(t)) - 2bv_5(t) - b^2 v_2(t) \\
\text{where, } \mathrm{Sigm}(v) &= \frac{5.00}{1 + \exp\{0.56(6.00 - v)\}}
\end{aligned}
\tag{1}
$$

where variables v_0, v_1, and v_3 represent the postsynaptic potential of pyramidal cells, excitatory interneurons, and inhibitory interneurons, respectively. Parameters A and B represent the E and I synaptic gains. Parameters a and b represent the inverse of time constant for E and I postsynaptic potential responses, respectively. $p(t)$ and C_n (where, $n = 1, ..., 4$) are external input and connectivity constants. In this study, C_n was fixed as $C_1 = 135$, $C_2 = 108$, and $C_3 = C_4 = 33.75$, drawing from previous studies [9, 24].

State and parameter estimation of the NM model

To track the time-varying changes in the NM model states and parameters, we applied an EnKF scheme. The EnKF is a well-known recursive data forecasting algorithm for data assimilation and is used in the research field of weather forecasting [11, 12]. Moreover, the EnKF works well for predicting time series of nonlinear dynamical systems. Since the observed EEG signal can be formulated using a nonlinear system such as the NM model, the EnKF scheme can be applied to directly estimate the system's state and

parameters in the NM model from the observed EEG data. However, the extent of the noise contained in the time series of observed neurophysiological signals such as EEG is usually unknown, and the noise component can be nonstationary. In addition, some parameters of the NM model contain the positive constraint in the range of its value (see Supplementary information for details). To estimate the state and parameters of the NM model from observed data while considering these issues, we combined a variational Bayesian noise adaptive algorithm in a linear/nonlinear Kalman Filter [14–18] and constrained state estimation in the Kalman Filter scheme [25–28] with a conventional EnKF [11, 12]. This method is hereinafter referred to as the variational Bayesian noise adaptive constrained EnKF (vbcEnKF). In this section, we will give a short description of the vbcEnKF algorithm for state and parameter estimation in the NM model. A more detailed mathematical description is given in the Supplementary information.

To apply the vbcEnKF approach to estimate the states and parameters in the NM model, the six first-order differential equations in the model were transformed to the state-space form, as shown below.

$$\dot{\mathbf{v}}_t = \begin{bmatrix} \dot{v}_0(t) \\ \dot{v}_1(t) \\ \dot{v}_2(t) \\ \dot{v}_3(t) \\ \dot{v}_4(t) \\ \dot{v}_5(t) \end{bmatrix} = \begin{bmatrix} v_3(t) \\ v_4(t) \\ v_5(t) \\ Aa\mathrm{Sigm}(v_1(t) - v_2(t)) - 2av_3(t) - a^2 v_0(t) \\ Aa[p(t) + C_2\mathrm{Sigm}(C_1 v_0(t))] - 2av_4(t) - a^2 v_1(t) \\ BbC_4\mathrm{Sigm}(C_3 v_0(t)) - 2bv_5(t) - b^2 v_2(t) \end{bmatrix} = g(\mathbf{v}_t, \theta_{\mathbf{t}}) \tag{2}$$

$$\text{state model}:$$

$$\begin{aligned} x_{t+1} &= \begin{bmatrix} G(\mathbf{v}_t, \theta_\mathbf{t}) \\ \theta_\mathbf{t} \end{bmatrix} + \zeta_t \\ &= f(\mathbf{v}_t, \theta_\mathbf{t}) + \zeta_t \\ &\text{where, } \zeta_t \sim \mathcal{N}(0, Q) \end{aligned} \tag{3}$$

$$\text{observation model}:$$

$$\begin{aligned} y_t &= Hx_t + w_t \\ &= [[0,\ 1,\ -1,\ 0,\ 0,\ 0],\ \mathbb{O}^{1\times 5}] \begin{bmatrix} \mathbf{v}_t \\ \theta_\mathbf{t} \end{bmatrix} + w_t \\ &\text{where, } \zeta_t \sim \mathcal{N}(0, \eta^{-1}R),\ \ \eta \sim \mathcal{G}(a_t, b_t) \end{aligned} \tag{4}$$

where $v_t = \{v_0(t),\ v_1(t),\ ....,\ v_5(t)\}$ and $\theta = \{A, a, B, b, p\}$ indicate the state and parameter vectors of the NM model, respectively. Note that $G(\mathbf{v}_t, \theta_t) = \int g(\mathbf{v}_t, \theta_t)dt$ was implemented with the fourth-order Runge-Kutta method in this study. ζ_t is state noise that follows the normal density $\mathcal{N}(0, Q)$. w_t is observation noise, which follows a normal distribution $\mathcal{N}(0, \eta^{-1}R)$. Moreover, η is a noise scale parameter that follows the gamma probability $\mathcal{G}(a_t, b_t)$. As mentioned above, applying the variational Bayesian scheme, the observation noise covariance $\eta^{-1}R$ is adaptively optimized by estimating the noise scale parameter, η, on a sample-by-sample basis [14–18]. See the Supplementary information for more details.

By applying the above-mentioned state-space form of the NM model to the vbcEnKF algorithm, model state and parameters were recursively

estimated from observed EEG data on a sample-by-sample basis. The whole algorithm of this method includes the following five steps:

1. Initialize the parameters for the probabilities $x_0 \sim \mathcal{N}(m_0, P_0)$, $\zeta_t \sim N(0, Q)$, $w_t \sim \mathcal{N}(0, \eta^{-1}R)$, and $\eta \sim \mathcal{G}(a_0, b_0)$.
2. Predict the system's state based on the current parameters of the probability density $\mathcal{N}(m_t, P_t)$ in the state model (see the Prediction step in Algorithm 1)
3. Get new observed data y_{obs_t} and update the parameters of probability densities of the state model and noise scaling factor η (see the Update step in Algorithm 1).
4. After updating the state x_{t+1}, if the constraints for variable x_{t+1} are violated, the projection of the updated variable x_{t+1} in the Update step (see Algorithm 1) is corrected so that the variables are within a feasible range.
5. Repeat steps 2–4.

For step 1, the initial parameter value of the probability density for the state variable x_t was set at $m_0 = \mathbb{O}^{1\times 11}$ (Note: $x_t = \{\mathbf{v}_t, \theta_t\}$, $\mathbf{v}_t \in \mathbb{R}^6$, $\theta_t \in \mathbb{R}^5$). The observation noise covariance R was set at $R = 50$ in both the numerical simulation and real EEG data analysis. The initial setting for $\mathcal{G}(a_0, b_0)$ was set to $a_0 = 1$ and $b_0 = 0.5$. The state noise covariance Q was selected differently depending on the analysis. In the numerical simulation, the covariance Q was fixed as $Q = \mathrm{diag}\left(\left[\Delta t \cdot 10^{-2} \cdot \mathbb{I}^{1\times 6}, 10^{-3} \cdot \mathbb{I}^{1\times 5}\right]\right)$. In the EEG analysis, Q was fixed as $Q = \mathrm{diag}\left(\left[\Delta t \cdot \mathbb{I}^{1\times 6}, 10^{-3} \cdot \mathbb{I}^{1\times 5}\right]\right)$. Δt indicates that the sampling interval was set as $\Delta t = 0.01$ in this study. The number of ensembles N_{ens} for the EnKF scheme was specified in each analysis (see the following section for more details). The summary of the whole algorithm of vbcEnKF is shown in Algorithm 1, and mathematical details are presented in the Supplementary information.

Algorithm 1 The variational Bayesian noise adaptive constrained Ensemble Kalman Filter (vbcEnKF) algorithm

(1) *Prediction step*:

$$\mathcal{X}_t^{(i)} = f(\mathcal{X}_{t-1}^{(i)}) + v_i, \text{ with } \mathcal{X}_{t-1}^{(i)} \sim \mathcal{N}(x_{t-1}, P_{t-1}) \text{ and } v_i \sim \mathcal{N}(0, Q)$$

$$x_t = \frac{1}{N_{ens}} \sum_i^{N_{ens}} \mathcal{X}_t^{(i)}, \text{ where, } N_{ens}: \text{ ensemble size.}$$

$$P_t = \frac{1}{N_{ens}-1} \sum_i^{N_{ens}} \left\{ (\mathcal{X}_t^{(i)} - x_t)(\mathcal{X}_t^{(i)} - x_t)^{\mathrm{T}} \right\} + Q$$

(2) *Update step*:

$$a_t = a_{t-1} + \frac{N}{2},\ b_t = b_{t-1},\ \langle \eta \rangle_\eta = a_t / b_t$$

$$\mathcal{Y}_t = \frac{1}{N_{ens}} \sum_i^{N_{ens}} \mathcal{Y}_t^{(i)}, \quad \text{with } \mathcal{Y}_t^{(i)} = H \cdot \mathcal{X}_t^{(i)}$$

$$P_{\mathcal{Y}\mathcal{Y}} = \frac{1}{N_{ens}-1} \sum_i^{N_{ens}} \left\{ (\mathcal{Y}_t^{(i)} - \mathcal{Y}_t)(\mathcal{Y}_t^{(i)} - \mathcal{Y}_t)^{\mathrm{T}} \right\} + \langle \eta \rangle_\eta^{-1} R$$

$$P_{\mathcal{X}\mathcal{Y}} = \frac{1}{N_{ens}-1} \sum_i^{N_{ens}} \left\{ (\mathcal{X}_t^{(i)} - x_t)(\mathcal{Y}_t^{(i)} - \mathcal{Y}_t)^{\mathrm{T}} \right\}$$

$$\text{KalmanGain}: \ K = P_{\mathcal{X}\mathcal{Y}} P_{\mathcal{Y}\mathcal{Y}}^{-1}$$

$$\mathcal{X}_{t+1}^{(i)} = \mathcal{X}_t^{(i)} + K(y_{obs_t} + w_i - \mathcal{Y}_t^{(i)}), \quad \text{with } w_i \sim \mathcal{N}\left(0, \langle \eta \rangle_\eta^{-1} R\right)$$

$$x_{t+1} = \frac{1}{N_{ens}} \sum_i^{N_{ens}} \mathcal{X}_{t+1}^{(i)}$$

$$P_{t+1} = P_t - K P_{\mathcal{Y}\mathcal{Y}} K^{\mathrm{T}}$$

$$b_{t+1} = b_t + \frac{1}{2} \frac{\|y_{obs_t} - H x_t\|^2 + H^{\mathrm{T}} P_t H}{R}$$

(3) *Constraint*:

$$\max: \ln P(x \mid y) \Longrightarrow \min: (x - x_{t+1})^{\mathrm{T}} P_{t+1}^{-1} (x - x_{t+1})$$

$$\text{such that, } d_L \le D x_{t+1} \le d_U$$

$$\hat{x}_L = x_{t+1} - P_{t+1} D^{\mathrm{T}} (D P_{t+1} D^{\mathrm{T}})^{-1} (D x_{t+1} - d_L)$$

$$\hat{x}_U = x_{t+1} - P_{t+1} D^{\mathrm{T}} (D P_{t+1} D^{\mathrm{T}})^{-1} (D x_{t+1} - d_U)$$

$$\text{if } d_{L,n} > x_{t+1,n} \ : \ x_{t+1,n} = \hat{x}_{L,n}$$

$$\text{if } d_{U,n} < x_{t+1,n} \ : \ x_{t+1,n} = \hat{x}_{U,n} \quad \text{where, } n = \{1, ..., N_s\},\ N_s : \# \text{ state}$$

Estimation of the mE/I ratio

To track time-varying changes in the E/I ratio based on the estimated model parameter regarding E and I synaptic gain, we proposed an index of E/I ratio named the model-based E/I ratio (mE/I ratio), which was calculated using the following equation:

$$\mathrm{mE/I}(t) = \frac{\hat{A}(t)}{\hat{A}(t) + \hat{B}(t)} \tag{5}$$

where $\hat{A}(t)$ and $\hat{B}(t)$ are E and I synaptic gain parameters estimated by our proposed method. The value of the mE/I ratio is sequentially calculated when updating the parameters in the NM model, on a sample-by-sample basis.

Numerical simulation

To confirm the validity of our proposed method, we first applied our method to synthetic data generated by the NM model to clarify whether our method can sequentially estimate changes in the model parameters. Moreover, since the EnKF algorithm is based on the sMC method for Bayesian probability estimation, we considered the following two conditions under which to evaluate the effects of state and parameter estimation accuracy. First, we tested the effects of the ensemble size for the sMC method in the EnKF scheme. In this simulation, the number of ensembles selected ranged from 40 to 500, with 20 steps. Second, we tested the effects of the initial seed of the random generator in the sMC. Since this value would affect the time evolution of the state estimation in the NM model, our proposed method was applied to the synthetic EEG data 50 times with different initial random seed values for each ensemble size condition.

The synthetic EEG were generated by the NM model using a numerical integration with the fourth order Runge-Kutta method with the time step $dt = 0.01$ (i.e., a sampling frequency of 100 Hz). The total sample length of the generated data was $3,000$ (i.e., the total time length of each dataset was $t = 30$ s). The generated synthetic EEG data consisted of two segments separated by the event of the parameter changes. The exact value of parameters and the timing of the parameter changes are shown below.

$$
\begin{aligned}
A(t) &= \begin{cases} 3.25 & (t \leq 15\text{s}) \\ 4.25 & (t > 15\text{s}) \end{cases} \\
a(t) &= 100 \\
B(t) &= \begin{cases} 22 & (t \leq 15\text{s}) \\ 19 & (t > 15\text{s}) \end{cases} \\
b(t) &= \begin{cases} 50 & (t \leq 15\text{s}) \\ 52 & (t > 15\text{s}) \end{cases} \\
p(t) &\sim \mathcal{N}(220, 22)
\end{aligned}
\tag{6}
$$

After generating synthetic EEG data with the above parameter settings, we applied our proposed method 50 times with a different initial random seed for each ensemble size condition. To test whether our method could correctly detect the parameter changes from observed data with noisy sampling, white noise was added to the synthetic EEG data using $\mathcal{N}(0, 1.3)$.

Next, we evaluated the estimation error score of synthetic EEGs and the standard deviation of estimated noise covariance, which was obtained from 50 trials of the estimation for each ensemble size condition. The estimation error score was calculated using the mean absolute error function described

by the following equation:

$$\mathrm{MAE} = \frac{1}{N_t}\sum_{t=0}^{N_t}|y_{obs,t} - \hat{y}_t|, \tag{7}$$

where $y_{obs,t}$ and $\hat{y}_t$ are the observed (exact) and predicted EEG at time sample t, and N_t is the total number of samples.

EEG analysis

To confirm its neurophysiological validity, we applied our proposed method to an open EEG dataset [13] to assess whether it could detect changes in E/I balance between the NREM/REM sleep stages.

Open EEG dataset

The dataset used was EEG data that were recorded with channels located at Pz-Oz based on a 10-10 system (sampling frequency = 100 Hz) in 197 participants' own homes during day-night periods by Kemp et al. (2000) [13]. A more detailed description of the task settings is provided in the original paper. This dataset is available at the following PysioNet repository: https://physionet.org/content/sleep-edfx/1.0.0/#ref1. During this validation, datasets that contained awake periods during the first NREM period in sleep marker data were excluded. As a result, we only used EEG data of 19 healthy participants who displayed a typical transition between sleep states during the first NREM and REM periods. Moreover, in the following validation, we used EEG data epochs that included the first 20-minute period from sleep onset time to the end of the first REM period for each participant.

Preprocessing of EEG

To reduce the effects of artifacts (e.g., nose movement, eye movement, and eye blink signals), we first applied an artifact removal method with empirical mode decomposition [29, 30]. A detailed description of this method is provided in the original articles [29, 30]. After removing artifacts, EEG data were band-pass filtered between 0.6 and 20 Hz with a zero-phase finite impulse response filter (the number of taps = 6,000; transition with 0.1 Hz; window function = Hanning window).

E/I ratio evaluation

After preprocessing, the EEG data were applied the vbcEnKF algorithm, and the mE/I ratio was sequentially calculated when updating the state and parameters of the NM model on a sample-by-sample basis. For the statistical analysis of the mE/I ratio, a temporally averaged mE/I ratio for the distinct five sleep stages (awake, stage 1, Stage 2, Stage 3/4, and REM) for each participant was calculated. The Kruskal-Wallis test was used to examine the differences in the mE/I ratio between the sleep stages. Moreover, to determine whether there were significant differences in the rank mean between specific pairs of sleep-stage-averaged mE/I ratios, Dunn's multiple comparisons test (two-tailed test) was also applied after the Kruskal-Wallis test.

Supplementary information. The Supplementary methods associated with this article can be found in the online version.

Acknowledgments. This research was partially funded by the JSPS KAKENHI grant (20K19867, https://kaken.nii.ac.jp/en/grant/KAKENHI-PROJECT-20K19867/) from the Japan Society for the Promotion of Science, and the support for in-house research on data science approaches to physiological research at National Institute for Physiological Sciences, Japan.

Competing interests. The authors declare no competing interests.

Data availability. The raw sleep EEG dataset obtained by Kemp et al. (2000) [13] is available at https://physionet.org/content/sleep-edfx/1.0.0/#ref1. All data used to reconstruct the results described in this paper are available at Github: https://github.com/myGit-YokoyamaHiroshi/EEG_vbcEnKF_mEI_est.

Code availability. The programming code implemented by Python language to reconstruct all results described in this paper is available at Github: https://github.com/myGit-YokoyamaHiroshi/EEG_vbcEnKF_mEI_est.

Author contributions. **Hiroshi Yokoyama**: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing original draft, Writing – review & editing. **Keiichi Kitajo**: Project administration, Supervision, Resources, Validation, Writing – review & editing.

References

[1] Yizhar, O. *et al.* Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature* **477** (7363), 171–178 (2011). https://doi.org/10.1038/nature10360 .

[2] Masuda, F. *et al.* Motor cortex excitability and inhibitory imbalance in autism spectrum disorder assessed with transcranial magnetic stimulation: a systematic review. *Translational Psychiatry* **9** (1) (2019). https://doi.org/10.1038/s41398-019-0444-3 .

[3] Li, X. *et al.* Tms-eeg research to elucidate the pathophysiological neural bases in patients with schizophrenia: A systematic review. *Journal of Personalized Medicine* **11** (5) (2021). https://doi.org/10.3390/jpm11050388 .

[4] Bridi, M. C. *et al.* Daily Oscillation of the Excitation-Inhibition Balance in Visual Cortical Circuits. *Neuron* **105** (4), 621–629.e4 (2020). URL https://doi.org/10.1016/j.neuron.2019.11.011. https://doi.org/10.1016/j.neuron.2019.11.011 .

[5] Tamaki, M. *et al.* Complementary contributions of non-REM and REM sleep to visual learning. *Nature Neuroscience* **23** (9), 1150–1156 (2020). URL http://dx.doi.org/10.1038/s41593-020-0666-y. https://doi.org/10.1038/s41593-020-0666-y .

[6] Stagg, C. J. *et al.* Local GABA concentration is related to network-level resting functional connectivity. *eLife* **2014** (3), 1–9 (2014). https://doi.org/10.7554/eLife.01465 .

[7] Premoli, I. *et al.* TMS-EEG signatures of GABAergic neurotransmission in the human cortex. *Journal of Neuroscience* **34** (16), 5603–5612 (2014). https://doi.org/10.1523/JNEUROSCI.5089-13.2014 .

[8] Belardinelli, P. *et al.* TMS-EEG signatures of glutamatergic neurotransmission in human cortex. *Scientific Reports* **11** (1), 1–14 (2021). URL https://doi.org/10.1038/s41598-021-87533-z. https://doi.org/10.1038/s41598-021-87533-z .

[9] Jansen, B. H. & Rit, V. G. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics* **73** (4), 357–366 (1995). https://doi.org/10.1007/BF00199471 .

[10] Kuhlmann, L. *et al.* Neural mass model-based tracking of anesthetic brain states. *NeuroImage* **133**, 438–456 (2016). URL http://dx.doi.org/10.1016/j.neuroimage.2016.03.039. https://doi.org/10.1016/j.neuroimage.2016.03.039 .

[11] Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* **99** (C5) (1994). https://doi.org/10.1029/94jc00572 .

[12] Houtekamer, P. L. & Mitchell, H. L. Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review* **126** (3), 796–811 (1998). URL http://journals.ametsoc.org/doi/10.1175/1520-0493(1998)126%3C0796:DAUAEK%3E2.0.CO;2. https://doi.org/10.1175/1520-0493(1998)126⟨0796:DAUAEK⟩2.0.CO;2 .

[13] Kemp, B., Zwinderman, A. H., Tuk, B., Kamphuisen, H. A. & Oberyé, J. J. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering* **47** (9), 1185–1194 (2000). https://doi.org/10.1109/10.867928 .

[14] Dong, Z. & Song, T. Variational inference of Kalman filter and its application in wireless sensor networks. *International Journal of Distributed Sensor Networks* **2013** (2013). https://doi.org/10.1155/2013/106434 .

[15] Sarkka, S. & Nummenmaa, A. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Transactions on Automatic Control* **54** (3), 596–600 (2009). https://doi.org/10.1109/TAC.2008.2008348 .

[16] Stroud, J. R. & Bengtsson, T. Sequential state and variance estimation within the ensemble Kalman filter. *Monthly Weather Review* **135** (9), 3194–3208 (2007). https://doi.org/10.1175/MWR3460.1 .

[17] Wang, G. Y. & Guan, B. L. Fuzzy adaptive variational Bayesian unscented Kalman filter. *Journal of Information Hiding and Multimedia Signal Processing* **6** (4), 740–749 (2015) .

[18] Wang, S. Y., Yin, C., Duan, S. K. & Wang, L. D. A Modified Variational Bayesian Noise Adaptive Kalman Filter. *Circuits, Systems, and Signal Processing* **36** (10), 4260–4277 (2017). https://doi.org/10.1007/s00034-017-0497-6 .

[19] Gao, R., Peterson, E. J. & Voytek, B. Inferring synaptic excitation/inhibition balance from field potentials. *NeuroImage* **158** (June), 70–78

(2017). URL http://dx.doi.org/10.1016/j.neuroimage.2017.06.078. https://doi.org/10.1016/j.neuroimage.2017.06.078 .

[20] Lendner, J. D. *et al.* An electrophysiological marker of arousal level in humans. *eLife* **9**, 1–29 (2020). https://doi.org/10.7554/eLife.55092 .

[21] Ullah, G. & Schiff, S. J. Assimilating seizure dynamics. *PLoS Computational Biology* **6** (5), 1–12 (2010). https://doi.org/10.1371/journal.pcbi.1000776 .

[22] Rahmati, V., Kirmse, K., Marković, D., Holthoff, K. & Kiebel, S. J. Inferring Neuronal Dynamics from Calcium Imaging Data Using Biophysical Models and Bayesian Inference. *PLOS Computational Biology* **12** (2), e1004736 (2016). URL https://dx.plos.org/10.1371/journal.pcbi.1004736. https://doi.org/10.1371/journal.pcbi.1004736 .

[23] Moye, M. J. & Diekman, C. O. Data Assimilation Methods for Neuronal State and Parameter Estimation. *Journal of Mathematical Neuroscience* **8** (1), 1–38 (2018). URL http://dx.doi.org/10.1186/s13408-018-0066-8. https://doi.org/10.1186/s13408-018-0066-8 .

[24] David, O. & Friston, K. J. A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage* **20** (3), 1743–1755 (2003). https://doi.org/10.1016/j.neuroimage.2003.07.015 .

[25] Luzar, M., Czajkowski, A., Witczak, M. & Korbicz, J. Actuators and sensors fault diagnosis with dynamic, state-space neural networks. *2012 17th International Conference on Methods and Models in Automation and Robotics, MMAR 2012* (August), 196–201 (2012). https://doi.org/10.1109/MMAR.2012.6347889 .

[26] Simon, D. & Chia, T. L. Kalman filtering with state equality constraints. *IEEE Transactions on Aerospace and Electronic Systems* **38** (1), 128–136 (2002). https://doi.org/10.1109/7.993234 .

[27] Simon, D. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Applications* **4** (8), 1303–1318 (2010). URL https://engagedscholarship.csuohio.edu/enece_facpubhttps://digital-library.theiet.org/content/journals/10.1049/iet-cta.2009.0032. https://doi.org/10.1049/iet-cta.2009.0032 .

[28] Wang, D., Chen, Y. & Cai, X. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research* **45** (11), 1–13 (2009). https://doi.org/10.1029/2008WR007401 .

[29] Flandrin, P., Goncalves, P. & Rilling, G. *Detrending and denoising with empirical mode decompositions*, Vol. 06-10-Sept (2004).

[30] Flandrin, P., Gonçalvès, P. & Rilling, G. in *EMD equivalent filter banks, from interpretation to applications* , Vol. 5 of *Interdisciplinary Mathematical Sciences* 57–74 (WORLD SCIENTIFIC, 2005). URL https://www.worldscientific.com/worldscibooks/10.1142/5862http://www.worldscientific.com/doi/abs/10.1142/9789812703347{_}0003.

Supplementary information:

A data assimilation method to track time-varying changes in the excitation-inhibition balance using scalp EEG

Hiroshi Yokoyama[1, 2, 3*], Keiichi Kitajo[1, 2*]

1. Division of Neural Dynamics, Department of System Neuroscience, National Institute for Physiological Sciences, 38 Nishigonaka, Myodaiji, Okazaki, Aichi, 444-8585, JAPAN
2. Department of Physiological Sciences, School of Life Science, Graduate University for Advanced Studies (SOKENDAI), 38 Nishigonaka, Myodaiji, Okazaki, Aichi, 444–8585, Japan
3. Program of Mathematical and Life Sciences, Graduate School of Integrated Sciences for Life, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima City, Hiroshima, 739-8526, JAPAN

*Corresponding author(s). E-mail(s): yokoyama@nips.ac.jp or hyokoyama@hiroshima-u.ac.jp ; kkitajo@nips.ac.jp

S1. *Variational inference for the Kalman Filter*

As stated in the main manuscript, to consider the effects of unknown or nonstationary observation noise, we applied a variational Bayesian noise adaptive approach for the Kalman Filter (vbKF; Dong & Song, 2013; Sarkka & Nummenmaa, 2009; Stroud & Bengtsson, 2007; Wang & Guan, 2015; Wang et al., 2017). First, we will explain the mathematical details of the noise adaptive algorithm. For the sake of simplicity, we will here give a detailed description of the algorithm for the vbKF in a Gaussian linear case. A discrete-time linear state space model is defined using the following equations:

$$x_{t+1} = f(x_t) + \zeta_t = Fx_t + \zeta_t, \quad \zeta_t \sim \mathcal{N}(0,\ Q),$$
$$y_t = h(x_t) + w_t = Hx_t + w_t, \quad w_t \sim \mathcal{N}(0, \eta^{-1}R), \eta \sim \mathcal{G}(a_t, b_t).$$

We here assumed that the probability of the observation noise follows the normal-gamma probability density. Furthermore, observation y_t and state variable x_t follow the multivariate Gaussian distribution $\mathcal{N}(Hx_t, \eta^{-1}R)$ and $\mathcal{N}(m_t, P_t)$, respectively. In such a case, if the state variable x_t and noise scaling parameter η are independent, the joint posterior probability density for these variables can be estimated with the variational Bayesian approximate, as below (Sarkka & Nummenmaa, 2009):

$$p(x_t, \eta | y_{1:t}) = p(y_t, x_t, \eta | y_{1:t-1}) \approx q(x_t)q(\eta).$$

The variational Bayesian approximation can be solved by minimizing the Kullback–Leibler (KL) divergence between the approximated distribution $q(x_t)q(\eta)$ and the exact posterior $p(y_t, x_t, \eta | y_{1:t-1})$, as below:

$$\mathrm{KL}\big(q(x_t)q(\eta)\ ||\ p(y_t, x_t, \eta | y_{1:t-1})\big) = \int \left\{ q(x_t)q(\eta) \times \ln\left(\frac{q(x_t)q(\eta)}{p(y_t, x_t, \eta | y_{1:t-1})} \right) \right\} dx_t d\eta.$$

Minimizing the right term of the above equation with respect to the probability $q(x_t)$, the solution of updating rule of the probability density for x_t is given as the following equation:

$$\begin{aligned}
\ln q(x_t) &= E\langle \ln p(y_t,\ x_t, \eta | y_{1:t-1}) \rangle_\eta + const. \\
&= E\langle \ln\{p(y_t\ | x_t, \eta)\ p(x_t |\ y_{1:t-1})\} \rangle_\eta + const. \\
&= -\frac{1}{2} x_t^T \left(P_t^{-1} + E\langle \eta \rangle_\eta^{-1} R^{-1} H^T H\right) x_t - \left(E\langle \eta \rangle_\eta R^{-1} y_t^T H + m_t^T P_t^{-1}\right) x_t + const. \\
&= \ln \mathcal{N}(x_t | m_{new}, P_{new}) + const.
\end{aligned}$$

This means that the parameters m_{new} and P_{new} can be calculated as follows:

$$P_{new} = \left(P_t^{-1} + \langle \eta \rangle_\eta R^{-1} H^T H\right)^{-1} \qquad = P_t - P_t H^T \left(\langle \eta \rangle_\eta^{-1} R + H P_t H^T\right)^{-1} H P_t,$$
$$m_{new} = P_{new}^{-1} \left(\langle \eta \rangle_\eta R^{-1} y_t^T H + m_t^T P_t^{-1}\right) = m_t + P H^T \left(\langle \eta \rangle_\eta^{-1} R + H P H^T\right)^{-1} (x_t - H m_t^T).$$

Note that $\langle \cdot \rangle_\eta$ indicates the expectation operator under the variable η.

In the same manner, the solution of $q(\eta)$ is also given, as follows:

$$\begin{aligned}
ln\, q(\eta) &= E\langle \ln p(y_t,\ x_t, \eta | y_{1:t-1}) \rangle_{x_t} + const. \\
&= E\langle \ln\{p(y_t\ | x_t, \eta)\ p(\eta | y_{t-1})\} \rangle_{x_t} + const. \\
&= E\left\langle \left(a_t + \frac{N}{2} - 1\right) \ln|\eta| - \left[b_t + \frac{1}{2}(y_t - Hx_t)^T R^{-1} (y_t - Hx_t)\right] \eta \right\rangle_{x_t} + const. \\
&= \ln \mathcal{G}(\eta | a_{new}, b_{new}) + const.
\end{aligned}$$

Therefore, the parameters a_{new} and b_{new} can be calculated as follows:

$$a_{new} = a_t + \frac{N}{2} \ ,$$

$$b_{new} = b_t + \frac{1}{2}\langle (y_t - Hx_t)^T R^{-1}(y_t - Hx_t)\rangle_{x_t} = b_t + \frac{1}{2}\ \frac{\left\{\|y_t - H^T m_t\|^2 + H^T P_t H\right\}}{R} \ .$$

The expectation of parameter η (i.e., $\langle\eta\rangle_\eta$) can be described as follows:

$$\langle\eta\rangle_\eta = \frac{a_{new}}{b_{new}}.$$

S2. *Inequality constraint for Kalman Filter*

In this section, we will explain the details of the state estimation in KF with an inequality constraint. In the context of the study on the control theory in robotics, various approaches of imposing the constraint in state estimation for the KF scheme have been proposed (Simon & Chia, 2002; Simon, 2010). Among these, we chose to apply the estimation projection approach with inequality constraints (Luzar et al., 2012; Simon & Chia, 2002; Simon, 2010; Wang et al., 2009) in our proposed method. In this approach, if the state variables x_t follow the probability density $x_t \sim p(x_t|x_{t-t}) = \mathcal{N}(m_t, P_t)$ with the constraint $Dx_t \leq d_U$, the KF algorithm estimates the value of m_t and P_t so as to solve the following problem:

$$\underset{x}{\text{argmax}}: \ln p(x_t|x_{t-1}) \ \Rightarrow \ \text{argmin} :(x_t - m_t)P_t^{-1}(x_t - m_t),$$

$$\text{s.t.,} \ \ Dx_t \leq d_U.$$

This problem can be solved using a Lagrange multiplier method, as follows:

$$L = (x_t - m_t)P_t^{-1}(x_t - m_t) + 2\lambda^T(Dx_t - d_U)$$

The solution is given as follows:

$$x_t = m_t - P_t D^T(DP_t D^T)^{-1}(Dm_t - d_U).$$

In the same manner, we can solve the solution in the case of the constraint $Dx_t \geq d_L$.

In the current study, to estimate the state and parameter using the EnKF scheme while considering some parameter constraints so as to appear the periodic signal in the NM model, we applied the following interval constraint:

$d_L \ \leq \ D\,x_t \leq \ d_U$, where $x_t = [v_t, \theta]^T = \left[[x_0, x_1, \ldots, x_5], [A, a, B, b, p]\right]^T$.

Note that parameters $D, d_L,$ and d_U followed with $D = [\mathbb{0}^{5\times6} \quad \mathbb{I}^{5\times5}]$, $d_L = [0.01, 5.00, 0.01, 5.00, 120]^T$, and $d_U = [100.00, 200.00, 100.00, 200.00, 320.00]^T$; that is, such interval constraints were set to restrict the range of values in the model parameters $A, a, B, b,$ and p. In particular, interval constraints for the parameter p in the NM model were selected so that the estimated value of p was satisfied with the interval $120 \leq p \leq 320$. This was chosen based on the study by Jansen and Rit (1995), which showed that the alpha-like periodic simulated signal was observed in the NM model when the external input p was distributed in the range of $120 \leq p \leq 320$. An interval constraint for the inverse of time constant a in excitatory post-synaptic potential (PSP) was chosen as $5 \leq a \leq 200$ to satisfy $5 \leq \tau_e[\text{ms}] \leq 200$, where $\tau_e = a^{-1}$. Also, the constraint for the inverse of time constant b in inhibitory PSP was

chosen to satisfy $5 \leq \tau_i[ms] \leq 200$, where $\tau_i = b^{-1}$. The parameters a and b were selected so that the NM model can generate delta to beta rhythms based on the report by David and Friston (2002).

S3. *Variational Bayesian noise adaptive constrained Ensemble Kalman Filter (vbcEnKF)*

S1 and S2 described the whole algorithm of the proposed vbcEnKF. To estimate state and model parameters in the NM model, we combined the variational Bayesian noise adaptive algorithm and inequality constraint approach with the EnKF. The whole vbcEnKF algorithm is summarized in the following table.

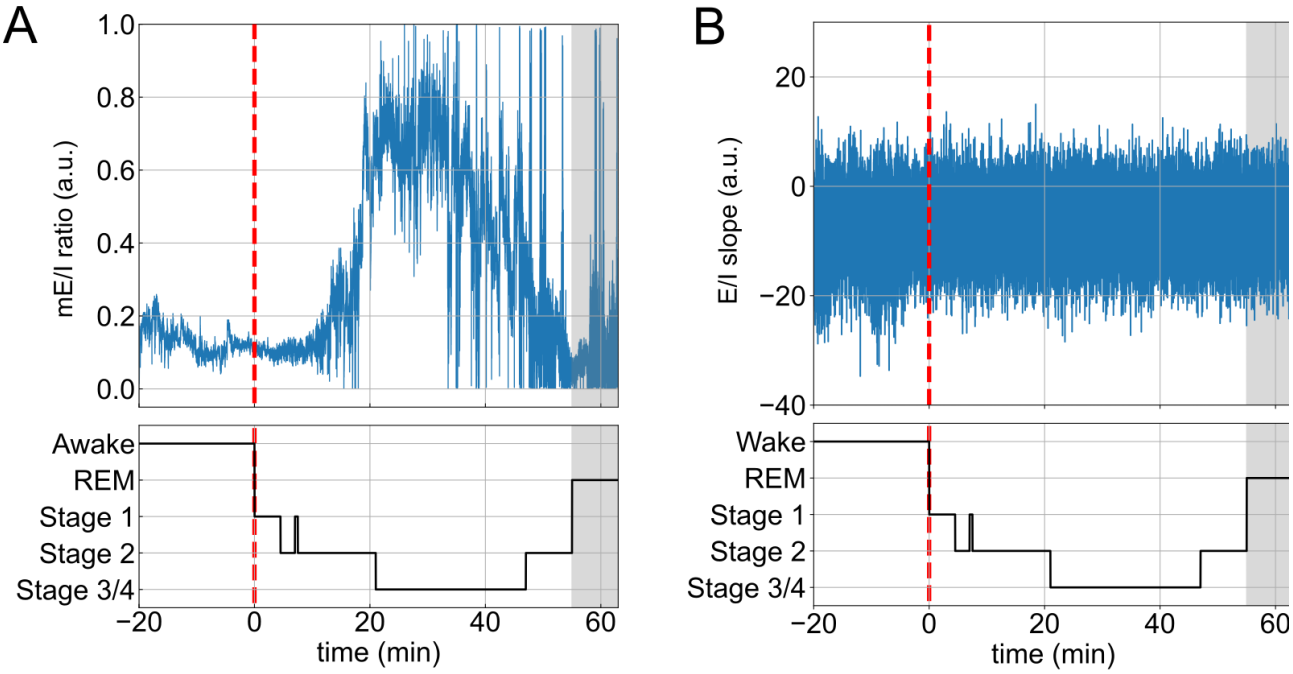
Supplementary Table 1: Algorithm of the vbcEnKF. The vbcEnKF consists of three steps. The first step is predicting the state variables and parameters in the NM model based on the current probability in the estimated state model. The second step is model updating the probability in the state and the observed model using observed data. The third step is applied only if the constraint for the updated state variables is violated.

1. Prediction step	$x_t^{(i)} = f\left(x_{t-1}^{(i)}\right) + v_t^{(i)}$ with $x_{t-1}^{(i)} \sim \mathcal{N}(x_{t-1}, P_{t-1})$ and $v_t^{(i)} \sim \mathcal{N}(0, Q)$ $x_t = \frac{1}{N_{ens}} \sum_i^{N_{ens}} x_t^{(i)}$, where N_{ens} is the ensemble size. $P_t = \frac{1}{N_{ens}-1} \sum_i^{N_{ens}} \left\{\left(x_t^{(i)} - x_t\right)\left(x_t^{(i)} - x_t\right)^T\right\} + Q$
2. Update step	$a_t = a_{t-1} + \frac{N}{2}, \quad b_t = b_{t-1}, \quad \langle\eta\rangle_\eta = a_t/b_t$ $\boldsymbol{y_t} = \frac{1}{N_{par}} \sum_i^{N_{par}} \boldsymbol{y_t^{(i)}}$ with $\boldsymbol{y_t^{(i)}} = H \cdot x_t^{(i)}$ $P_{yy} = \frac{1}{N_{par}-1} \sum_i^{N_{par}} \left\{\left(\boldsymbol{y_t^{(i)}} - \boldsymbol{y_t}\right)\left(\boldsymbol{y_t^{(i)}} - \boldsymbol{y_t}\right)^T\right\} + \langle\eta\rangle_\eta^{-1} R$ $P_{xy} = \frac{1}{N_{par}-1} \sum_i^{N_{par}} \left\{\left(x_t^{(i)} - x_t\right)\left(\boldsymbol{y_t^{(i)}} - \boldsymbol{y_t}\right)^T\right\}$ $K = P_{xy} P_{yy}^{-1}$ $x_{t+1} = \frac{1}{N_{par}} \sum_i^{N_{par}} \left\{\hat{x}_i + K\left(y_t + w_i - \boldsymbol{y_t^{(i)}}\right)\right\}$ with $w_i \sim \mathcal{N}\left(0, \langle\eta\rangle_\eta^{-1} R\right)$ $P_{t+1} = P_t - K P_{yy} K^T$ $b_{t+1} = b_t + \frac{1}{2} \frac{\lVert y_t - H \cdot x_t \rVert^2 + H^T P_t H}{R}$
3. Constraint	$\min : (x - x_{t+1}) P_{t+1}^{-1} (x - x_{t+1})$ such that, $d_L \leq D x_{t+1} \leq d_U$ $\hat{x}_L = x_{t+1} - P_{t+1} D^T (D P_{t+1} D^T)^{-1} (D x_{t+1} - d_L)$ $\hat{x}_U = x_{t+1} - P_{t+1} D^T (D P_{t+1} D^T)^{-1} (D x_{t+1} - d_U)$ if $d_{L,n} > x_{t+1,n}$: $x_{t+1,n} = \hat{x}_{L,n}$ if $d_{U,n} < x_{t+1,n}$: $x_{t+1,n} = \hat{x}_{U,n}$, where, $n = \{1, \ldots, N_s\}$, N_s: # state

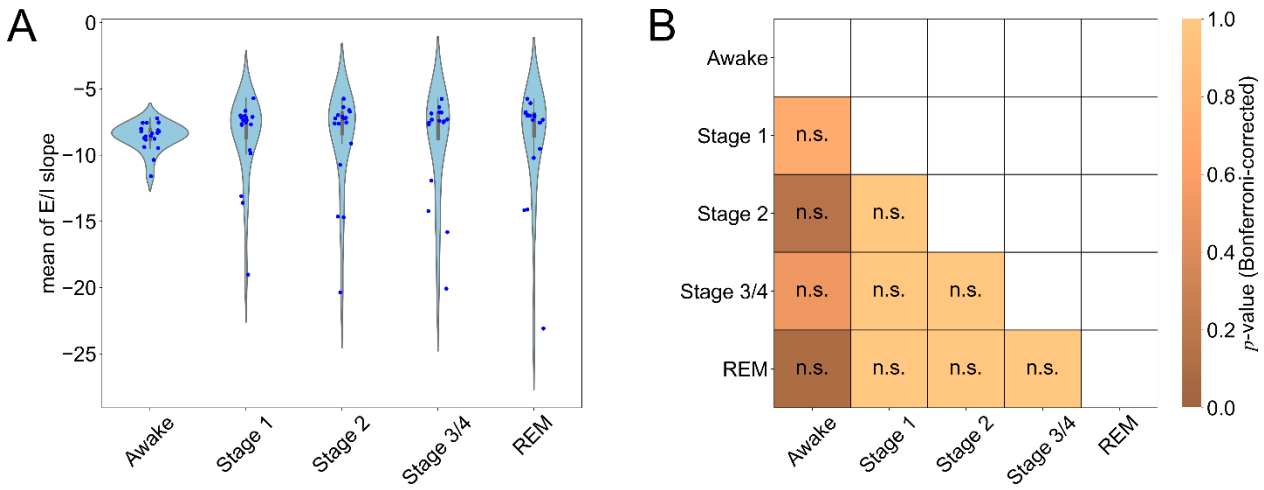
S4. *Supplementary analysis of sleep EEG*

As stated in the Discussion section of the main manuscript, to compare the estimation results of sleep-dependent changes in the E/I ratio between our proposed and prior methods, we applied the previous method of E/I estimation proposed by Gao et al. (2017) to the same sleep EEG datasets (Kemp et al., 2000) that we used in this study. The previous E/I estimation method proposed by Gao et al. (2017) evaluated the feature of E/I balance changes based on power spectral density (PSD) analysis in EEG signals. As mentioned in the Discussion section of the main manuscript, Gao et al. (2017) found that the slope of the log-log plot of PSD from 30 to 50 Hz (i.e., 1/f power law exponent in the gamma frequency band) was correlated with the changes in the E/I ratio based on numerical simulation; this indicates that the feature of E/I balance can be quantified based on this slope (hereinafter called the E/I slope method). More details on the E/I slope method can be found in the original article (Gao et al., 2017). The results of applying the E/I slope method to the sleep EEG datasets used to test our own method are shown in Supplementary Figures 1 and 2.

In Supplementary Figure 1, we present a typical example of the comparison of the results in E/I ratio estimation between the two methods (our method and the E/I slope method) using the same EEG data (participant ID: 4032). As shown in this figure, while the mE/I ratio estimated by our proposed method changes according to the progress in the sleep stage, the E/I slope estimated by the prior method (Gao et al., 2017) could not identify sleep-dependent changes. Moreover, the group-level analysis of E/I slope data (see Supplementary Figure 2) revealed no significant difference in the E/I slope between sleep stages. As mentioned in the main manuscript, the inability of the E/I slope method to detect sleep-dependent changes in EEG could be a result of the effects of artifacts and measurement noise. In general, gamma band oscillations of human scalp EEG tend to be compromised due to electromyogenic artifacts from cranial and ocular muscles. Therefore, the results shown in Supplementary Figures 1 and 2 suggested that PSD-based analysis, such as the E/I slope method, is directly affected by the signal quality of observed EEG for the E/I balance estimation.



Supplementary Figure 1: Comparison of our proposed method with a prior method (Gao et al., 2017). A) Time-varying changes in the mE/I ratio estimated using our method (participant ID: 4032). B) Result of time-varying changes in the E/I slope estimated by the prior method (Gao et al., 2017, participant ID: 4032). Blue lines indicate the estimated E/I scores. Lower panels in A) and B) show the time profiles of sleep stage changes in participant 4032. Red dotted lines indicate sleep onset (onset of the NREM stage). Gray shaded areas indicate the REM period.



Supplementary Figure 2: Sleep-state dependent changes in the E/I slope using Gao et al.'s model (2017). A) The violin plot and box plot were obtained from the samples of the averaged E/I slope in 19 participants for each sleep stage. The box and error bar of the boxplot included in the violin plot present the interquartile range and 95% confidence intervals, respectively. The middle line of the box shows the median value of the samples. The blue dots indicate samples of the averaged E/I slope for each participant. B) Multiple comparisons of pairs of the averaged E/I slope between sleep stage. Dunn's multiple comparison test was applied after the Kruskal-Wallis test.

Supplementary References

David, O. & Friston, K. J. A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *NeuroImage* **20** (3), 1743–1755 (2003). https: //doi.org/10.1016/j.neuroimage.2003.07.015 .

Dong, Z. & Song, T. Variational inference of Kalman filter and its application in wireless sensor networks. *International Journal of Distributed Sensor Networks* **2013** (2013). https://doi.org/10.1155/2013/106434 .

Gao, R., Peterson, E. J. & Voytek, B. Inferring synaptic excitation/inhibition balance from field potentials. *NeuroImage* **158** (June), 70–78 (2017). URL http://dx.doi.org/10.1016/j.neuroimage.2017.06.078. https: //doi.org/10.1016/j.neuroimage.2017.06.078 .

Jansen, B. H. & Rit, V. G. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics* **73** (4), 357–366 (1995). https://doi.org/10.1007/ BF00199471 .

Luzar, M., Czajkowski, A., Witczak, M. & Korbicz, J. Actuators and sensors fault diagnosis with dynamic, state-space neural networks. *2012 17th International Conference on Methods and Models in Automation and Robotics, MMAR 2012* (August), 196–201 (2012). https://doi.org/10. 1109/MMAR.2012.6347889 .

Sarkka, S. & Nummenmaa, A. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Transactions on Automatic Control* **54** (3), 596–600 (2009). https://doi.org/10.1109/TAC. 2008.2008348 .

Simon, D. & Chia, T. L. Kalman filtering with state equality constraints. *IEEE Transactions on Aerospace and Electronic Systems* **38** (1), 128–136 (2002). https://doi.org/10.1109/7.993234 .

Simon, D. Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Applications* **4** (8), 1303–1318 (2010). https://doi.org/10.1049/iet-cta.2009.0032 .

Stroud, J. R. & Bengtsson, T. Sequential state and variance estimation within the ensemble Kalman filter. *Monthly Weather Review* **135** (9), 3194–3208 (2007). https://doi.org/10.1175/MWR3460.1 .

Wang, D., Chen, Y. & Cai, X. State and parameter estimation of hydrologic models using the constrained ensemble Kalman filter. *Water Resources Research* **45** (11), 1–13 (2009). https://doi.org/10.1029/ 2008WR007401 .

Wang, G. Y. & Guan, B. L. Fuzzy adaptive variational Bayesian unscented Kalman filter. *Journal of Information Hiding and Multimedia Signal Processing* **6** (4), 740–749 (2015) .

Wang, S. Y., Yin, C., Duan, S. K. & Wang, L. D. A Modified Variational Bayesian Noise Adaptive Kalman Filter. *Circuits, Systems, and Signal Processing* **36** (10), 4260–4277 (2017). https://doi.org/10.1007/ s00034-017-0497-6 .