Inferring a simple mechanism for alpha-blocking by fitting a neural population model to EEG spectra

Agus Hartoyo[1*], Peter J. Cadusch[2], David T. J. Liley[3,4], Damien G. Hicks[1,2,5†]

1 Centre for Micro-Photonics, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
2 Department of Physics and Astronomy, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
3 Centre for Human Psychopharmacology, School of Health Sciences, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
4 Department of Medicine, University of Melbourne, Parkville, Victoria 3010, Australia
5 Bioinformatics Division, Walter & Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia

*ahartoyo@swin.edu.au; †dghicks@swin.edu.au

Abstract

Alpha blocking, a phenomenon where the alpha rhythm is reduced by attention to a visual, auditory, tactile or cognitive stimulus, is one of the most prominent features of human electroencephalography (EEG) signals. Here we identify a simple physiological mechanism by which opening of the eyes causes attenuation of the alpha rhythm. We fit a neural population model to EEG spectra from 82 subjects, each showing different degrees of alpha blocking upon opening of their eyes. Although it is notoriously difficult to estimate parameters from fitting such models, we show that, by regularizing the differences in parameter estimates between eyes-closed and eyes-open states, we can reduce the uncertainties in these differences without significantly compromising fit quality. From this emerges a parsimonious explanation for the spectral changes between states: Just a single parameter, p_{ei}, corresponding to the strength of a tonic, excitatory input to the inhibitory population, is sufficient to explain the reduction in alpha rhythm upon opening of the eyes. When comparing parameter estimates across different subjects we find that the inferred differential change in p_{ei} for each subject increases monotonically with the degree of alpha blocking observed. In contrast, other parameters show weak or negligible differential changes that do not scale with the degree of alpha attenuation in each subject. Thus most of the variation in alpha blocking across subjects can be attributed to the strength of a tonic afferent signal to the inhibitory cortical population.

Author summary

One of the most striking features of the human electroencephalogram (EEG) is the presence of neural oscillations in the range of 8-13 Hz. It is well known that attenuation of these alpha oscillations, a process known as alpha blocking, arises from opening of the eyes, though the cause has remained obscure. In this study we infer the mechanism underlying alpha blocking by fitting a neural population model to EEG spectra from 82 different individuals. Although such models have long held the promise of being able to relate macroscopic recordings of brain activity to microscopic neural parameters, their

utility has been limited by the difficulty of inferring these parameters from fits to data. Our approach is to fit both eyes-open and eyes-closed EEG spectra together, minimizing the number of parameter changes required to transition from one spectrum to the other. Surprisingly, we find that there is just one parameter, the external input to the inhibitory neurons in cortex, that is responsible for attenuating the alpha oscillations. We demonstrate how the strength of this inhibitory input scales monotonically with the degree of alpha blocking observed over all 82 subjects.

Introduction

Alpha blocking is a classic feature of the human electroencephalogram (EEG). First identified by Hans Berger as part of his discovery of human EEG in the 1920's [1, 2], it is now arguably its most robust empirical feature. Classically, alpha blocking refers to the reduction in spontaneously-recorded occipital alpha band (8-13 Hz) power in response to opening of the eyes [3]. More generally, changes in alpha-band power can be effected by a range of visual, tactile and auditory stimuli and altered states of arousal and is widely used as a diagnostic of cognitive activity [4–7].

Despite the importance of alpha blocking in studies of cognition, it still lacks a generally-accepted, mechanistic understanding [8]. Importantly, the mechanism associated with alpha blocking is usually considered separately from the mechanism associated with alpha wave generation. Whereas cortical alpha is thought to be generated by feedforward and feedback interactions between the thalamus and overlying cortex [9–12], blocking is considered to arise from changes in the phase synchrony of populations of these near-identical cortico-thalamic alpha oscillators [13–15]. In this paper, we show how alpha generation and blocking can be described self-consistently within in a single neural population model for the cortex.

Neural population models describe how microscopic properties in the cortex, such as post-synaptic rate constants, affect macroscopic observables, such as the local field potential detected by the EEG [16–18]. These models match the high time resolution and low spatial resolution of the EEG and have long been used to interpret the characteristics of alpha-band activity [12, 19, 20]. Notably, it has been shown how, with judiciously chosen model parameter values, alpha oscillations can arise spontaneously in the cortex [13, 21–26], without the need for direct pacing by oscillatory inputs [27–30]. However, it has been more difficult to interpret alpha blocking within these models since there are multiple ways to reduce or eliminate alpha activity [31]. For example, alpha attenuation has been attributed to coincident changes in several thalamo-cortical parameters controlling the feedforward, cortico-thalamo-cortical, and intra-cortical circuits [32].

A fundamental challenge in using neural population models is the difficulty in estimating parameter values directly from fits to EEG data [33]. Although forward calculations have provided plausible explanations for spontaneous alpha generation, solving the inverse problem to determine the many unknown model parameters is crucial if we want to relate the subject-to-subject variability observed in EEG signals to an associated variability in certain microscopic parameters. This will help identify (and potentially control) the underlying microscopic drivers of the EEG response, and associated cognitive behavior, in a given individual.

Recently [33], we quantified the large parameter uncertainties associated with fitting a neural population model [22, 26] to EEG data. Such large, and correlated, uncertainties are a consequence of the unidentifiability [34, 35] and sloppiness [36, 37] of the model and are typical when fitting many parameters. Our study found that only one out of the 22 parameters was individually identifiable. This single identifiable parameter, the decay rate of the inhibitory post-synaptic potential γ_i, was discovered by

fitting to EEG spectra exhibiting alpha oscillations from subjects with their eyes closed. The value of this inhibitory decay rate needed to be within a narrow range in order to generate alpha oscillations (of any amplitude) from a white noise input, regardless of the values of the other parameters. This showed the fundamental importance of this parameter in generating spontaneous alpha-band activity.

To understand alpha blocking we must confront another aspect of the unidentifiability problem: whether one can learn the *change* in a parameter in response to a particular stimulus, in this case the opening of the eyes. This is important since it is often more useful to know how much a parameter changes in response to a stimulus than it is to know the absolute value of that parameter before or after the stimulus is applied. We refer to this as the 2-state fitting problem since this will involve fitting two spectra (eyes closed and eyes open) from a single individual. Thus our previous study [33], where we only fit to the eyes-closed spectra in each individual, was a 1-state fitting problem.

Naively, it would seem that the unidentifiability we found for the 1-state problem would doom the 2-state fitting problem since one seemingly needs to perform separate fits to each state. However, fitting the two states simultaneously and by penalizing parameter differences between the states, we are able to reliably determine the change, or differential response, of a particular parameter, even though the absolute value of that parameter in each state can be quite uncertain. When examining data across many subjects, we are able to associate a single parameter p_{ei} - the strength of extra-cortical input to the inhibitory cortical population - with the attenuation of alpha oscillations upon opening of the eyes. This unifies the mechanisms for alpha generation and blocking within a single model.

In the rest of this introduction we briefly describe the data, the model, and the fitting strategy. Further details about methodology are given in the "Methods of analysis" section.

EEG data

The EEG data used in this study is provided in the online repository [38] (https://www.physionet.org/pn4/eegmmidb/). We use data from the occipital electrode from 82 individuals, as in our previous study [33], although this time we use eyes-open as well as eyes-closed data. We apply Welch's method [39] to estimate the 2×82 power spectra. Once again, because of the well-known nonlinearities and nonstationarities in EEG recordings, we restrict our study to frequencies between 2 Hz and 20 Hz. Since the absolute power in the EEG data is not meaningful, each spectrum is normalized to have a total power of 1. Our interest is thus in changes in spectral shape, not magnitude, upon going from eyes closed to eyes open states in each individual.

EEG data variability across individuals

It is well-known that the degree to which alpha rhythm is attenuated by an identical visual stimulus varies across individuals [40] with, for example, a negative correlation with age [41]. Visual inspection of the spectra we use from the 82 subjects shows that there is substantial variability in the degree of alpha blocking across individuals used in this study (see Fig 1 for a sample set, Fig A S1 Appendix for the full set).

Our approach is to use this individual variability to quantify how much each parameter shifts between EC to EO states and how these shifts scale with the degree of alpha blocking. To do this quantitatively, we needed to define a measure of alpha blocking strength. Here we use the Jensen-Shannon divergence, D_{JS}, which provides a scalar measure of the difference in shape between (normalized) eyes-closed (EC) and eyes-open (EO) spectra from each individual (see Section "Jensen-Shannon divergence

as a measure of the degree of alpha blocking"). To demonstrate how this measure aligns with our intuitive notion of spectrum change, the EC and EO data in Fig 1 are ordered by increasing D_{JS}. Although we explored alternative measures such as the change in relative strength of the alpha band component, we use D_{JS} since it is a more global measure of function change that does not rely on defining a particular frequency band. As we will show, by comparing how parameter differences scale with increasing D_{JS} we are able to establish how much each (microscopic) parameter changes with the degree of alpha-blocking.

Neural population model

The model used in this paper is the local variant of the mean-field model originally described in Refs [22, 26]. As described in our previous study [33], this model consists of 10 coupled non-linear ODEs parameterized by 22 physiologically-motivated parameters (see Table 1). Local equations are linearized around a fixed point and the power spectral density (PSD) is derived assuming a stochastic driving signal of the excitatory population that represents thalamo-cortical and long range cortico-cortical inputs, assumed to be Gaussian white noise. The modelled PSD can then be written as a rational function of frequency derived from the transfer function for the linearized system. As was explained in earlier studies [33, 42], tonic excitatory signals to the inhibitory (p_{ei}) and excitatory(p_{ee}) populations are included as unknown parameters to account for potential DC offsets in extracortical inputs.

In this study, we use the identical model but with two changes. The first is to introduce an additional parameter to allow for a non-white background spectrum (giving a total of 23 parameters - see Table 1). Though this adds an extra degree of freedom it is necessary in order to achieve fits to some of the eyes-open spectra. In fact there is evidence from EEG and ECoG studies (for example, see [43]) that the background PSD may have a frequency dependence (typically quoted as $1/f$) not readily accounted for by a rational transfer function alone. While various approaches have been suggested to account for such a dependence, we have chosen the simplest way to incorporate it into our model by relaxing the white noise assumption and using coloured noise for the driving signal. Specifically, we take the input PSD, $S^{\text{in}} \propto 1/f^{\eta}$ where η is the exponent of the input spectrum treated as a new state-dependent adjustable parameter in the range $0 \leq \eta \leq 2$; $\eta = 0$ corresponds to the original white noise, $\eta = 1$ to the pure $1/f$ (pink) noise, and $\eta = 2$ to a Wiener process (Brownian noise).

The second change is to incorporate the main result learned from [33] and restrict the range of γ_i. There it was found that the inhibitory rate constant γ_i has a sharply peaked posterior distribution, making it (uniquely) identifiable in the eyes-closed case. This was reproduced here in the eyes-open data when the EEG had a detectable peak in the alpha band; if no peak was observed, the posterior distribution resembled the assumed prior distribution. In light of this and in line with the search for a parsimonious explanation for alpha blocking, the prior distributions for the eyes-closed and eyes open cases in the current study were both limited to a reduced interval around the range found for its posterior distribution in [33] (See the updated minimum and maximum value for γ_i in Table 1).

Model fitting strategy

In this 2-state fitting problem, the EC spectrum and the EO counterpart from a given subject are treated as a single dataset to be jointly fit by the model. Given that a single spectrum fit has 23 unknown parameters, a naive fit to two spectra would have 46 potentially unknown parameters.
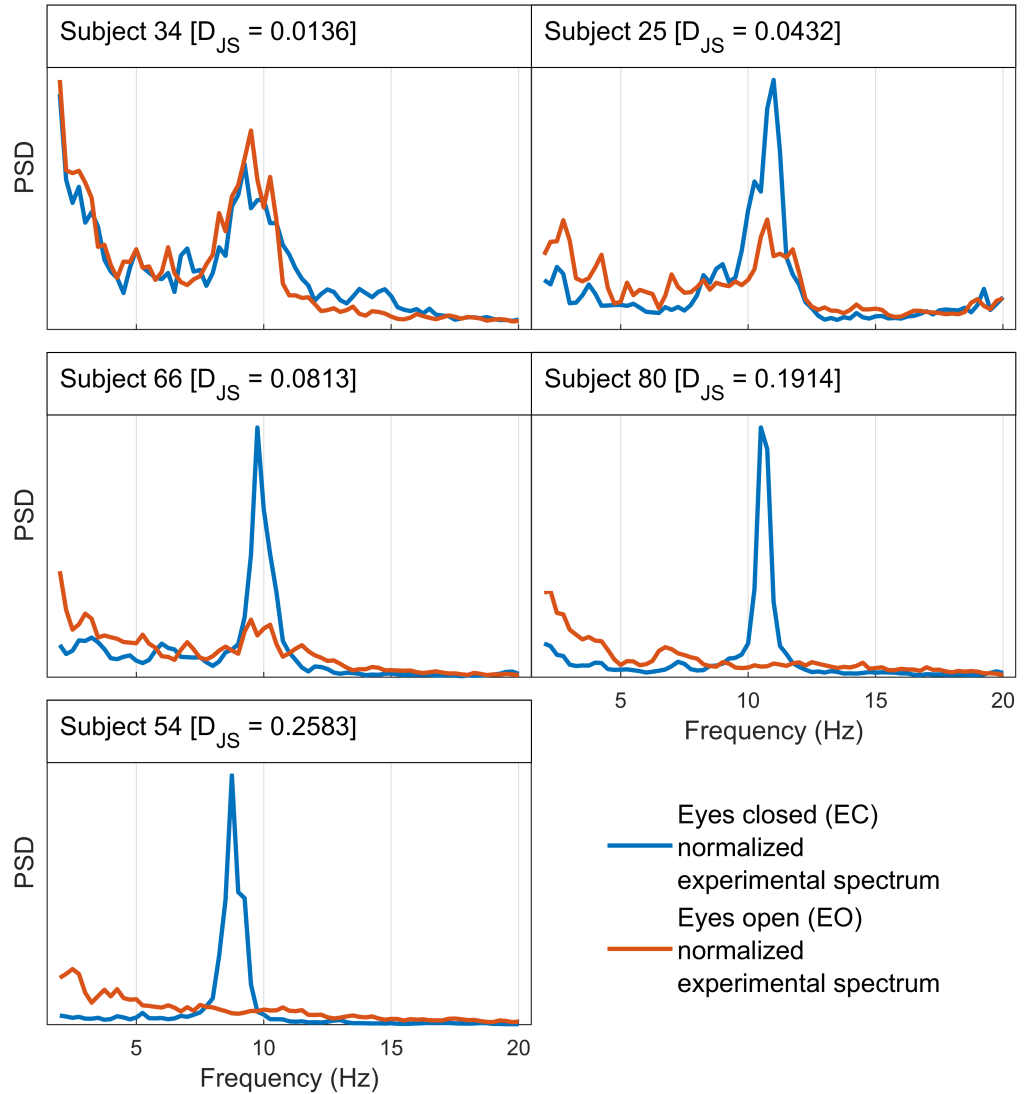
Fig 1. Different subjects exhibit different degrees of alpha blocking upon opening of the eyes. Here five subjects have been selected to illustrate the range of alpha blocking behaviour observed in the dataset. The vertical axis on each plot represents an arbitrary scale for the normalized power spectral density (PSD). Some subjects do not show any reduction in alpha power between EC and EO states (e.g. Subject 34); others exhibit partial blocking where the alpha activity in EO state is weaker than that of EC but is still pronounced (e.g. Subject 25); while some show total blocking where the alpha activity in the EO spectra completely disappears (e.g. Subject 80). To quantify the degree to which the EEG spectrum changes upon opening of the eyes, we compute the Jensen-Shannon divergence, D_{JS}, between the eyes-closed (EC) and eyes-open (EO) normalized experimental spectrum for each subject. A larger value of D_{JS} implies more pronounced EEG spectrum changes, or alpha-wave suppression. The complete set of spectra for all subjects is presented in Fig A in S1 Appendix, ordered by D_{JS}.

Table 1. State-distinct parameters and state-common parameters.

Type	No	Label	Description	Interval	No	Label
State-distinct parameters	1	τ_e	Passive membrane decay time const. of the excitatory population	[5, 150] ms	1	$\tau_e(EC)$
					2	$\tau_e(EO)$
	2	τ_i	Passive membrane decay time const. of the inhibitory population	[5, 150] ms	3	$\tau_i(EC)$
					4	$\tau_i(EO)$
	3	γ_e	Excitatory postsynaptic potential rate constant	[0.1, 1.0] /ms	5	$\gamma_e(EC)$
					6	$\gamma_e(EO)$
	4	γ_i	Inhibitory postsynaptic potential rate constant	[0.01, 0.1] /ms	7	$\gamma_i(EC)$
					8	$\gamma_i(EO)$
	5	Γ_e	Postsynaptic potential amplitude of the excitatory population	[0.1, 2.0] mV	9	$\Gamma_e(EC)$
					10	$\Gamma_e(EO)$
	6	Γ_i	Postsynaptic potential amplitude of the inhibitory population	[0.1, 2.0] mV	11	$\Gamma_i(EC)$
					12	$\Gamma_i(EO)$
	7	p_{ee}	Rate of the excitatory input to the excitatory population	[0.0, 10.0] /ms	13	$p_{ee}(EC)$
					14	$p_{ee}(EO)$
	8	p_{ei}	Rate of the excitatory input to the inhibitory population	[0.0, 10.0] /ms	15	$p_{ei}(EC)$
					16	$p_{ei}(EO)$
	9	η	Exponent of the input spectrum	[0.0, 2.0]	17	$\eta(EC)$
					18	$\eta(EO)$
State-common parameters	10	h_e^{rest}	Mean resting membrane potential of the excitatory population	[-80, -60] mV	19	$h_e^{rest}(EC, EO)$
	11	h_i^{rest}	Mean resting membrane potential of the inhibitory population	[-80, -60] mV	20	$h_i^{rest}(EC, EO)$
	12	h_e^{eq}	Mean Nernst membrane potential of the excitatory population	[-20, 10] mV	21	$h_e^{eq}(EC, EO)$
	13	h_i^{eq}	Mean Nernst membrane potential of the inhibitory population	[-90, -65] mV ‡	22	$h_i^{eq}(EC, EO)$
	14	$S_e^{\max}$	Maximum mean firing rate of the excitatory population	[0.05, 0.5] /ms	23	$S_e^{\max}(EC, EO)$
	15	$S_i^{\max}$	Maximum mean firing rate of the inhibitory population	[0.05, 0.5] /ms	24	$S_i^{\max}(EC, EO)$
	16	$\bar{\mu}_e$	Firing thresholds of the excitatory population	[-55, -40] mV	25	$\bar{\mu}_e(EC, EO)$
	17	$\bar{\mu}_i$	Firing thresholds of the inhibitory population	[-55, -40] mV	26	$\bar{\mu}_i(EC, EO)$
	18	σ_e	Std. deviation of firing thresholds of the excitatory population	[2, 7] mV	27	$\sigma_e(EC, EO)$
	19	σ_i	Std. deviation of firing thresholds of the inhibitory population	[2, 7] mV	28	$\sigma_i(EC, EO)$
	20	N_{ee}^{β}	# of connections an excitatory neuron receives from excitatory neurons	[2000, 5000]	29	$N_{ee}^{\beta}(EC, EO)$
	21	N_{ei}^{β}	# of connections an inhibitory neuron receives from excitatory neurons	[2000, 5000]	30	$N_{ei}^{\beta}(EC, EO)$
	22	N_{ie}^{β}	# of connections an excitatory neuron receives from inhibitory neurons	[100, 1000]	31	$N_{ie}^{\beta}(EC, EO)$
	23	N_{ii}^{β}	# of connections an inhibitory neuron receives from inhibitory neurons	[100, 1000]	32	$N_{ii}^{\beta}(EC, EO)$

The model is characterized by 23 physiological parameters associated with a given subject. As the subject moves from the EC state to the EO state, so do the physiological parameters. A *state-distinct parameter* is a physiological parameter that changes between states and corresponds to two distinct fitting-parameters. A *state-common parameter* is kept the same for both the EC and EO states and corresponds to a single fitting-parameter. There are 9 state-distinct parameters (which translate into twice as many fitting parameters) and 14 state-common parameters giving a total of 32 adjustable parameters to optimize during the joint fitting to both spectra for each individual. Minimum and maximum values for the physiological parameters are presented. The list of the physiologically-plausible intervals was originally proposed in [42], and is here updated with a reduced interval for γ_i as suggested by the identifiability analysis conducted in [33]. ‡The physiologically-plausible interval for h_i^{eq} presented in this table corrects a typographical error made in [33] which incorrectly indicated the parameter's minimum and maximum to be -20 mV and 10 mV, respectively.

To reduce the number of unknowns we implement two types of constraint. The first constraint(see Section "State-common parameters and state-distinct parameters") is that 14 of the parameters should remain the same in both the EC and EO conditions. This set is referred to as *state-common parameters*. The remaining 9 parameters are allowed to vary between conditions and are thus referred to as *state-distinct parameters*, giving a total of 32 unknown parameters. This joint-fitting approach for the two spectra allows us to couple together the dependency between the EC and EO parameters while at the same time allowing their actual values to be determined by the data.

The second constraint (see Section "Regularization of parameter differences") is to penalize (regularize) non-zero differences between EC and EO values for state-distinct parameters. This helps to identify the important parameter differences driving the change in spectral shape from EC to EO. Our regularization procedure is a variant on the standard procedure employed in high-dimensional inference problems searching for sparse, or parsimonious, solutions [44].

We use the same fitting scheme to that described in [33]: Fitted parameters are obtained using particle swarm optimization (PSO) [45, 46] starting from a random set of initial states. Each of the 82 subjects was fit separately as a parallel job on the OzStar supercomputer at Swinburne University of Technology, generating 1000 independent fit samples per subject. Computations were performed using a parallel for-loop with 30 workers and 30 CPUs each with 1 GB of memory. From the resulting sample of 1000 optimized parameter sets, the 10 percent with the lowest cost function values are accepted as final estimates (a detailed discussion justifying this threshold was given previously [33]).

Further details on data analysis are given in Section "Methods of analysis". Our implementation of the methods and all datasets are publicly available at https://github.com/cds-swinburne/Hartoyo-et-al-2020-DATA-n-CODE.

Results

Fig 2 shows the best model fits to EC and EO spectra from 5 different subjects, ordered vertically by degree of alpha blocking. Both regularized and unregularized cases exhibit good fits to the data. The similarity between regularized and unregularized cases confirms that the bias caused by regularization is within acceptable limits.

The EC and EO posterior marginal distributions for each parameter are shown in Fig 3. Plots for the 5 subjects are ordered vertically by degree of alpha blocking, as in Fig 2. Distributions are estimated from the 100 best fit parameter sets for each subject. p_{ei} shows the most noticeable difference between its EC and EO distributions, with EO distributions drifting increasingly higher than their corresponding EC distributions as alpha blocking gets larger. Differences between distributions for EC and EO states are weakly visible for p_{ee} and mostly negligible for other parameters.

To better quantify the difference between EC and EO states for each parameter and how it scales with the degree of alpha blocking, we calculate the difference between each EC to EO parameter estimate. We do this for each of the $N_J = 100$ best sample fits found for each of the $N_I = 82$ subjects. Thus, if θ_{ij} is a given parameter estimate indexed by subject, i, and sample fit, j, we define the *parameter response*, $\Delta\theta_{ij}$, to be

$$\Delta\theta_{ij} = \theta_{ij}^{\text{EO}} - \theta_{ij}^{\text{EC}}, \qquad i \in \{1, \ldots, N_I\}, \quad j \in \{1, \ldots, N_J\} \tag{1}$$

$$\overline{\Delta\theta_i} = \frac{1}{N_J} \sum_{j=1}^{N_J} \left(\theta_{ij}^{\text{EO}} - \theta_{ij}^{\text{EC}} \right) \tag{2}$$

where $\overline{\Delta\theta_i}$ is the resulting mean parameter response from EC to EO (averaged over sample fits) for a given subject, i.
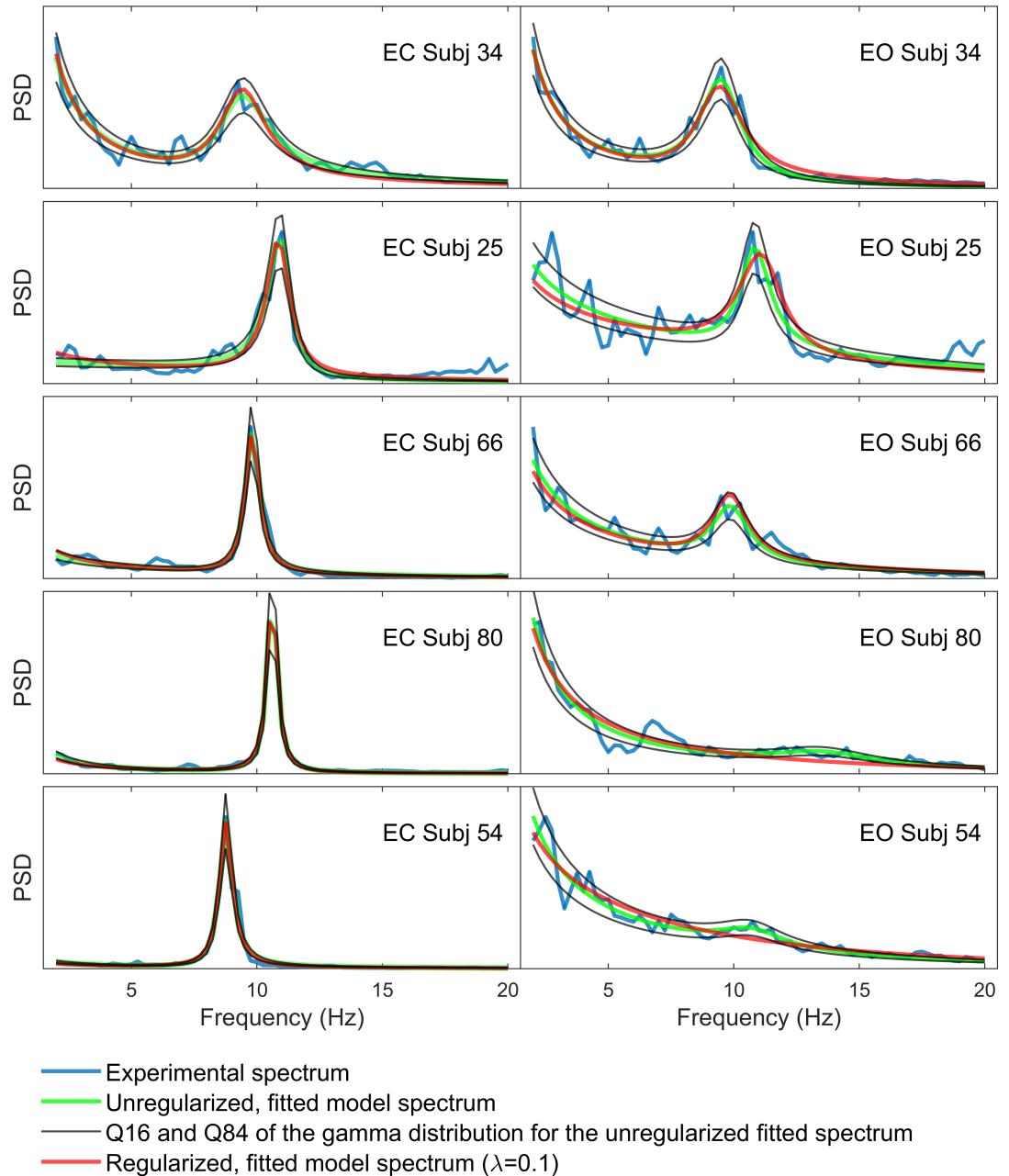
Fig 2. Regularized and unregularized best fits to EC and EO spectra. Best fit results for the 5 subjects shown in Fig 1. Subjects are ordered vertically by the degree of alpha blocking, with alpha blocking increasing downwards. Regularized fits (red) deviate only slightly from the unregularized fits (green). The 16% and 84% uncertainty quantiles (based on the gamma distribution for the unregularized best fits) are shown in black. These boundaries define the acceptable error of a fit. Regularized best fits deviate only slightly from the unregularized ones and generally stay within these uncertainty quantiles. In order to visualize the different fits, EC and EO spectra for a given subject are not necessarily shown on the same vertical scale.
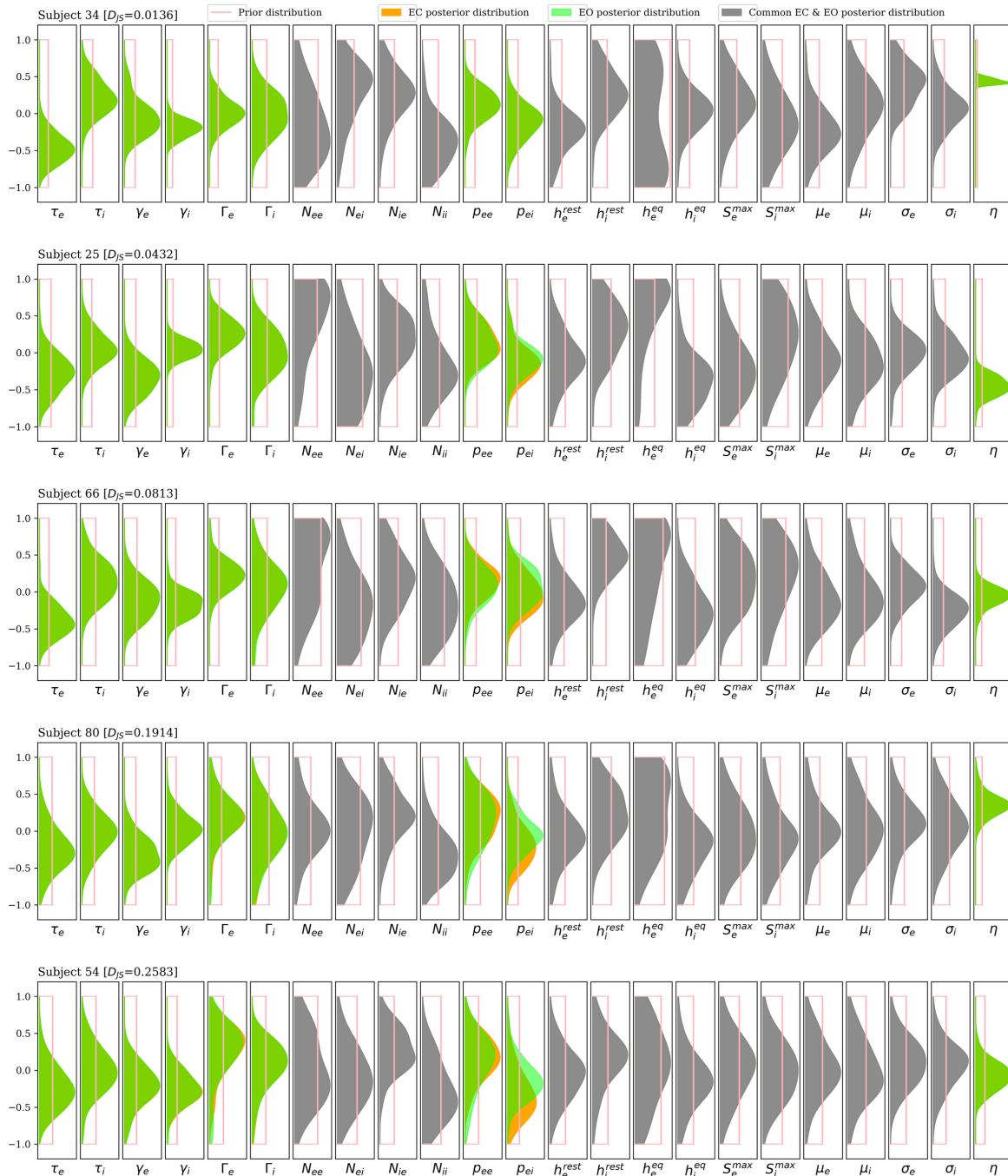
Fig 3. Posterior distributions for each parameter. Posterior distributions for state-distinct parameters (with EC in orange and EO in green) and state-common parameters (grey), again for the 5 subjects in Fig 1 and Fig 2. Subjects are ordered vertically by the degree of alpha blocking, with alpha blocking increasing downwards. The distributions are calculated using kernel density estimates from the best 100 of 1000 randomly seeded particle swarm optimizations for each subject. Each parameter is plotted in normalized coordinates, where -1 corresponds to the lower limit of the plausible parameter interval and +1 corresponds to the upper limit. The parameter p_{ei} is the only parameter where the difference between EC and EO distributions increases consistently with the degree of alpha blocking. Weaker shifts in p_{ee} are also apparent.

In Fig 4, to examine the association between each parameter response and the degree of alpha blocking, we plot $\overline{\Delta\theta_i}$ versus D^i_{JS} for each of the 82 subjects i. We also plot the 25% to 75% interquartile range determined from that subject's $N_J = 100$ sample fits, which provides an estimate of the unidentifiability of the parameter response. Results are shown for all 9 state-distinct parameters.

To characterize how each parameter response scales with the degree of alpha blocking, we perform a linear regression of $\Delta\theta$ versus D_{JS}. We use linearity simply to characterize the trend, not because of any expectation of linearity. Most parameter responses are either zero or show an insignificant trend with the degree of alpha blocking. The major exception is Δp_{ei} which increases monotonically with increasing D_{JS}. p_{ee} shows a non-zero parameter response, but its trend with D_{JS} is weak and restricted to low values. In the context of our model this implies that excitatory input to the inhibitory population is the dominant factor determining the response of alpha oscillations to a visual stimulus.

Discussion

By fitting a neural population model to EEG data from 82 individuals, we have demonstrated a clear association between the degree of alpha blocking and a single model parameter: the strength of a tonic excitatory input to the inhibitory population, p_{ei}. Most of the change between eyes-closed and eyes-open spectra is explained by variation in this external input level.

As a consistency check, we perform a forward calculation to test how the EEG spectrum is affected by each state-distinct parameter. In Fig 5 we compare the spectra calculated from the best fit parameter set (for a particular subject), to the spectra calculated when the best-fit values for the 9 state-distinct parameters are perturbed. Results show that the magnitude of the alpha rhythm is most sensitive to perturbations of p_{ei}, with increasing p_{ei} resulting in less alpha-band power. This is consistent with the tendency for p_{ei} to increase with alpha blocking (Fig 4). Interestingly, decreases in p_{ee} also cause a weaker alpha peak, although the effect is considerably less sensitive that for p_{ei}. We note that the relative effects of different parameter perturbations can vary among the different individuals, making it important to compare data across multiple individuals when performing the inverse problem.

The sensitivity of the alpha peak amplitude to changes in p_{ei} helps explain why the inverse problem identified p_{ei} appears as the dominant driver of alpha blocking: regularization is, after all, designed to identify sensitive input parameters. While this consistency is comforting, it does not rule out the role of other factors. One could, for example, contrive large changes in multiple weakly-sensitive parameters to give the same effect as a small change in a single, sensitive parameter. These are, in fact, the types of solution that a fit commonly finds without any regularization. Thus, in our effort to tame the unidentifiability problem, we are pushed towards simplicity as a guiding principle for identifying causes.

Although this model does not identify the source of the extra-cortical input, p_{ei}, we speculate that the visual stimulus increases thalamo-cortical afference to cortical neurons. This shifts the fixed point to a position in state space with a weaker alpha rhythm. Importantly, our model states that this extracortical input is tonic, rather than oscillatory. Our single-parameter explanation for the difference between eyes-closed and eyes-open spectra contrasts with previous explanations for alpha blocking which invoked changes in multiple parameters [31, 32].

Importantly, we have now shown how both alpha generation and blocking can arise within a single model in a way that is justified by fits to real EEG spectra. Our previous study [33] found that the presence of spontaneous alpha oscillations was
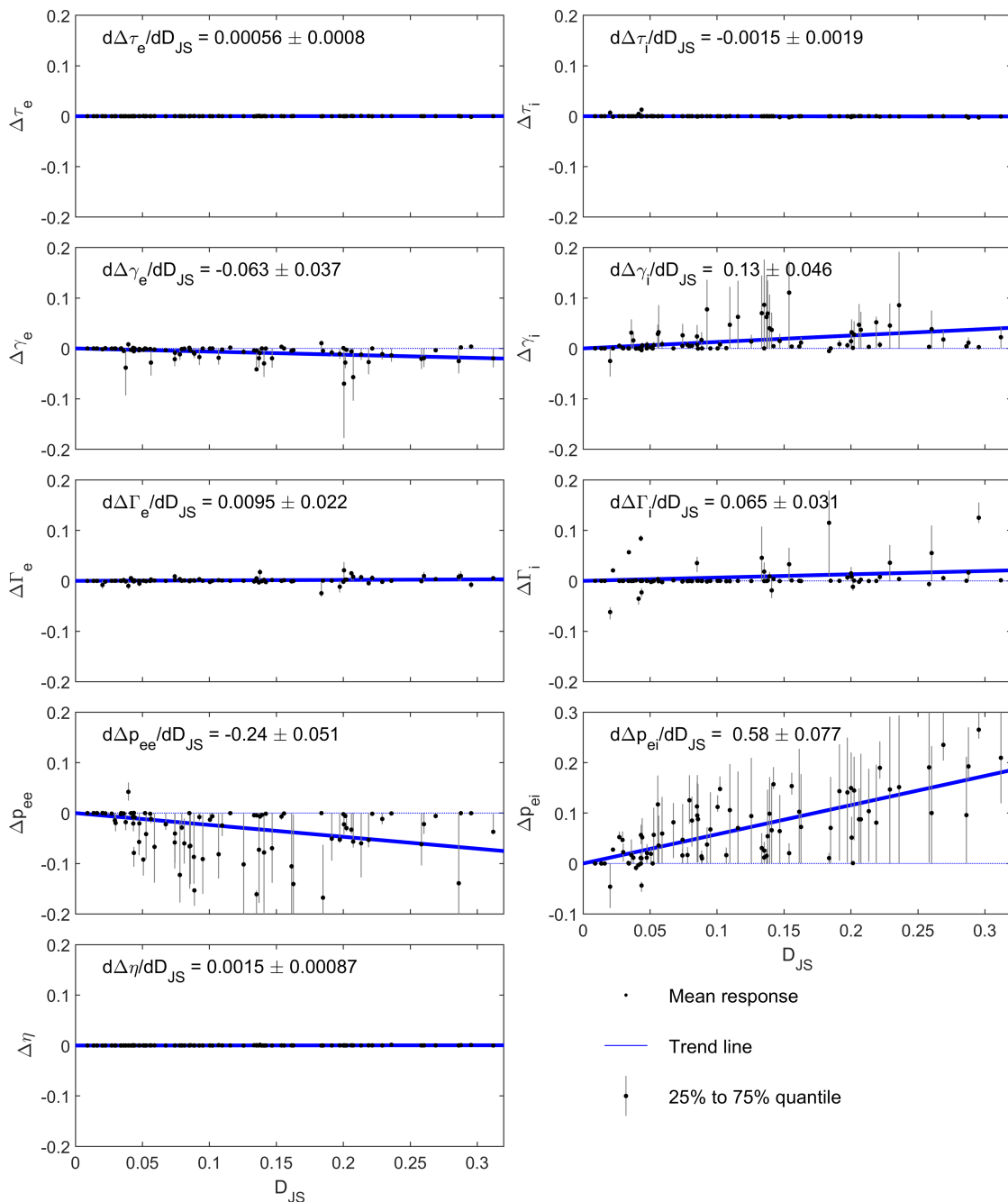
Fig 4. EC to EO parameter responses and how they scale with the degree of alpha blocking. The EC-to-EO parameter response (Eq 1) is calculated from the 100 best samples fits for each of the 82 subjects. The mean (black dot), calculated from Eq 2, and interquartile ranges (error bar) for each subject are plotted against the Jensen-Shannon divergence, D_{JS}, for that subject. In order to quantify how much each parameter response scales with the degree of alpha blocking we performed a linear regression through the sample fits; errors in the fit were estimated by randomly sampling from the distributions estimated from the sample fits. The resulting trend line is shown in blue, with its slope and error reported on each subplot. Several of the parameters ($\tau_e, \tau_i, \Gamma_e, \eta$) show essentially zero response to alpha blocking. Of the others, only Δp_{ei} (lower right subplot) shows a clear trend, increasing monotonically with D_{JS}. p_{ee} shows a non-zero parameter response but its trend with D_{JS} is weak and not monotonic. This result suggests that alpha blocking by visual stimulus can largely be attributed to an increase in a tonic afferent signal p_{ei} to the inhibitory cortical population, with weak or negligible contributions from the other parameters.
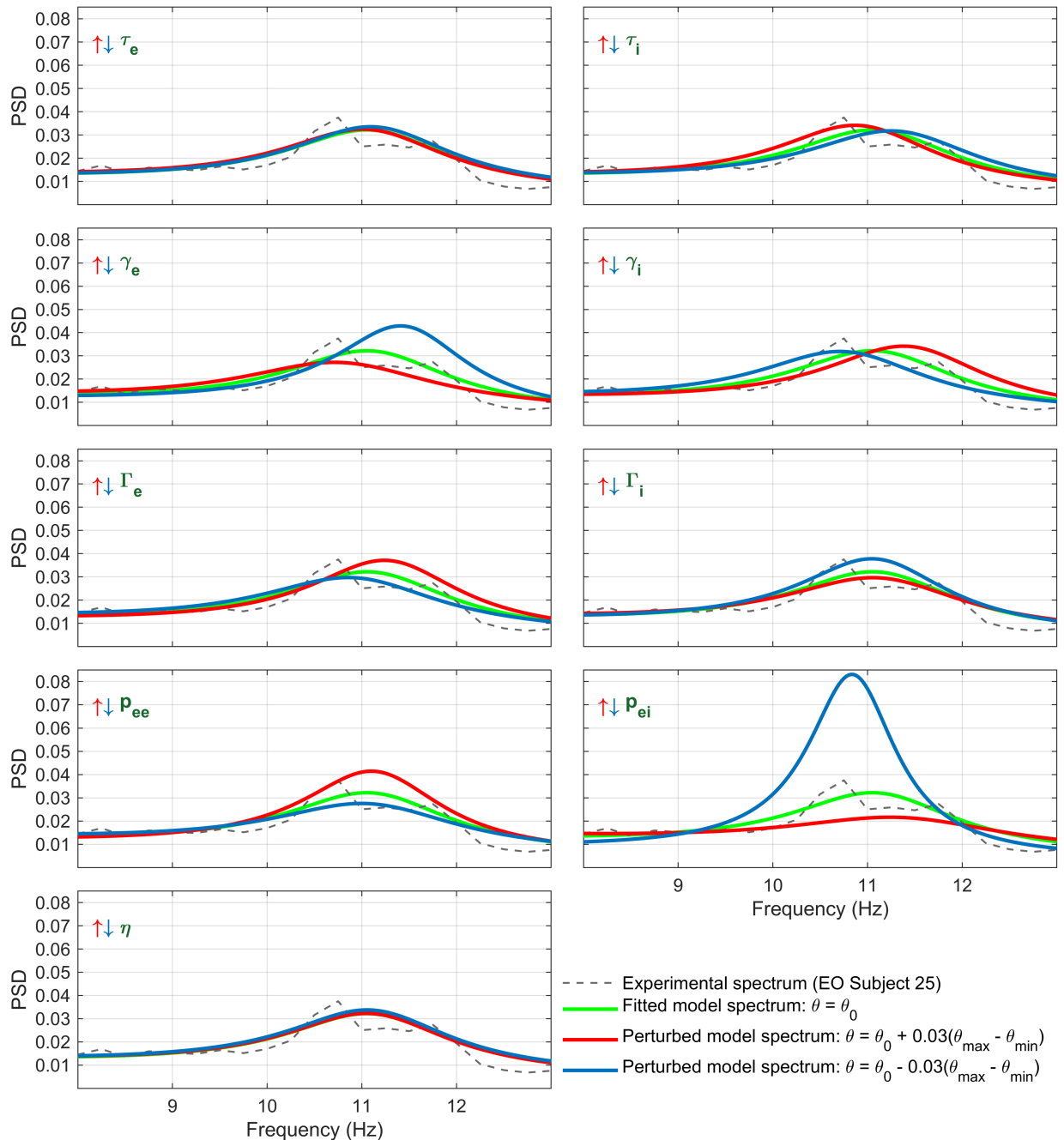
Fig 5. Forward calculation of the sensitivity of the alpha-rhythm to individual parameters. Shown are calculations depicting the sensitivity of the alpha-rhythm to each of the nine state-distinct parameters. The initial state (green) is that of the best fit for EO Subject 25. Each parameter is then perturbed by +3% (red) or -3% (blue) of the plausible interval, keeping other parameters constant. We observe that perturbing p_{ei} changes the alpha rhythm amplitude most significantly, with a comparatively small change to the peak frequency. The same perturbations applied to p_{ee} had a similar type of effect, though reversed and to a smaller extent. Alpha band power is only weakly affected by γ_e or γ_i though they both control the frequency. We note in general that perturbations applied to the other parameters have significantly smaller effects than perturbations to p_{ei}.

crucially dependent on the value of a single parameter - the decay rate of the inhibitory post-synaptic potential, γ_i. This demonstrated how intracortical inhibition is the essential driver of alpha wave activity. Our present work shows how extracortical input to the inhibitory cortical neurons, p_{ei}, is the driver of classical alpha blocking. Together we have thus identified the respective loci of physiological control for both the generation and attenuation of alpha oscillations.

In the future, the approach we have described could be used to determine the parameter response associated with anesthetic induction. Changes in EEG spectra under general anesthesia, from the loss of consciousness to the period of anaesthetic maintenance, are well characterized [47, 48]. Implementing the procedure we have outlined here, may allow us to identify a subset of the parameters driving the changes of brain state, connecting them to specific disruptions in interneuronal communication associated with a particular anesthetic. This will provide quantitative insight into the mechanisms underlying the loss of consciousness.

Methods of analysis

Jensen-Shannon divergence as a measure of the degree of alpha blocking

The Jensen-Shannon divergence, D_{JS}, is closely related to the Kullback-Leibler divergence [49, 50]. It is symmetric, non-negative, finite, and bounded [51]. D_{JS}, is traditionally used to measure the difference between two probability distributions. Here we use it to measure the difference between the EC and EO spectra for each subject since these spectra have the same properties as a probability distribution: they are non-negative with a total integral of 1 (since the spectra are normalized as described in Section "EEG data"). D_{JS} thus measures the difference in *shape* between EC and EO spectra, since power differences among original experimental spectra are irrelevant due to the normalization.

If the Kullback-Leibler divergence of $\boldsymbol{P}$ relative to $\boldsymbol{Q}$ is given by

$$D_{KL}(\boldsymbol{P}||\boldsymbol{Q}) = \int p(x) \ln \frac{p(x)}{q(x)} dx \tag{3}$$

the Jensen-Shannon divergence between the EC normalized experimental spectrum $\boldsymbol{S}^{EC}$ and the EO normalized experimental spectrum $\boldsymbol{S}^{EO}$ is given by

$$
\begin{aligned}
D_{JS}(\boldsymbol{S}^{EC}||\boldsymbol{S}^{EO}) = {} & \frac{1}{2} D_{KL}(\boldsymbol{S}^{EC}||\frac{1}{2}(\boldsymbol{S}^{EC} + \boldsymbol{S}^{EO})) \\
& + \frac{1}{2} D_{KL}(\boldsymbol{S}^{EO}||\frac{1}{2}(\boldsymbol{S}^{EC} + \boldsymbol{S}^{EO})).
\end{aligned}
\tag{4}
$$

In this work the logarithmic base e is used in the calculation of the Jensen-Shannon divergence, in which case $0 \leq D_{JS} \leq \ln 2$.

We have chosen the Jensen-Shannon divergence based on the postulate that the greater the degree of alpha blocking, the larger the D_{JS}. This is qualitatively confirmed by examination of the spectra from different subjects (Fig 1 and Fig A in S1 Appendix). We are interested how the parameter response scales with D_{JS} (and thus alpha blocking), since this relates changes in the model to changes in the spectra. To check this, Fig C in S1 Appendix shows how the total parameter response (given by the Manhattan distance in parameter space, $\sum_m |\Delta\theta_m|$, where m indexes state-distinct parameters) increases monotonically with D_{JS}. Distances in spectral space, captured by D_{JS}, thus scale smoothly with distances in parameter space, providing a link between microscopic parameters and macroscopic observable over 82 different subjects.

Physiological interpretability depends, of course, on whether *individual* parameters scale with D_{JS} (see Fig 4).

State-common parameters and state-distinct parameters

There are compelling physiological reasons why certain parameters should have the same value in EC and EO states in a single individual. The EEG signal transitions reversibly between the EC and EO state in times of the order of a second. It is unlikely that parameters determined largely by the morphology or connectivity of the neurons could vary significantly on this time-scale. We thus do not expect the average number of synapses per neuron (N_{ee}^{β}, N_{ei}^{β}, N_{ie}^{β}, and N_{ii}^{β}) to vary between the two states. Similarly parameters representing intrinsic neuronal properties such as those involved in the sigmoidal response of the neural population (maximum firing rates ($S_e^{\max}$, $S_i^{\max}$) and slope (σ_e, σ_i) and the threshold ($\bar{\mu}_e$, $\bar{\mu}_i$), resting (h_e^{rest}, h_i^{rest}) and equilibrium (h_e^{eq}, h_i^{eq}) potentials) plausibly could be expected to remain constant on this time scale. We thus require that these 14 parameters, referred to as state-common parameters, have the same value in EC and EO states for a particular individual. We emphasize that, although each of these parameters has a shared value across states, that value can vary between individuals.

The remaining 9 parameters are allowed to vary between states and are thus referred to as state-distinct parameters. Together with state-common parameters, this gives $14 + 2 \times 9 = 32$ distinct parameters down from the maximum possible $2 \times 23 = 46$ free parameters.

It is interesting to note that the state-distinct parameters fall into two sub-groups: those that characterize the input to the macro-column (i.e. tonic levels of p_{ee} and p_{ei} and the exponent of the input spectrum, η), and those that affect the shape, amplitude and time-scale of the post-synaptic potentials (γ_e, γ_i, Γ_e, Γ_i, τ_e, and τ_i). Some or all of the parameters in the state-distinct group could conceivably vary on such a time scale (though with different levels of plausibility) and so we allowed all of them to vary between states. However, after fitting and regularization, we discovered that it is primarily parameters from the first subgroup of state-distinct parameters (particularly p_{ei}, and to a lesser extent p_{ee}) that play the dominant role in distinguishing EC from EO spectra. The second sub-group are thus *a posteriori* shared parameters as distinct from the state-common parameters which are *a priori* shared.

Regularization of parameter differences

A straightforward least-squares fit of EC/EO pairs resulted in parameter differences between states shows little systematic scaling with the degree of alpha blocking (see Fig B in S1 Appendix). We hypothesized that this was caused by parameter unidentifiability (uncertainty) obscuring the subtle differences between states. To address this problem, we added a regularization term to our least-squares cost function.

Regularization is a standard method for identifying the subset of sensitive parameters in a fit by trading off fitting accuracy [44]. In traditional regularization, using for example the L1 norm [52], it is the value of the parameter itself that is regularized (penalized). In our case, we penalize the *differences* between (state-distinct) parameters, rather than the parameter values themselves, biasing most to zero and allowing only the most important ones to be non-zero. This reduces much of the unwanted variation caused by sloppy parameters.

The regularized cost function for the 32-parameter fit is given by

$$C = \frac{1}{2} \sum_n (\alpha \hat{S}_n(\boldsymbol{\theta}) S_n^{in} - S_n)^2_{EC} + \frac{1}{2} \sum_n (\alpha \hat{S}_n(\boldsymbol{\theta}) S_n^{in} - S_n)^2_{EO}$$
$$+ \frac{\lambda}{N_D} \sum_{\theta_m \in \mathcal{D}} |\hat{\theta}_m^{EO} - \hat{\theta}_m^{EC}| \tag{5}$$

where $\boldsymbol{\theta}$ is the 32-parameter vector to be optimized, $\alpha \hat{S}(\boldsymbol{\theta})$ is the model spectrum normalized by the scaling factor α (the formula to compute α appears as Eq (12) in [33]), S^{in} is the input spectrum given by $1/f^\eta$, S is the experimental spectrum, $\hat{\theta}_m$ is the parameter θ_m normalized to the range of [-1,1] corresponding to the θ_m's plausible range, $\mathcal{D}$ is the set of state-distinct parameters, N_D is the number of the state-distinct parameters, and λ is the regularization parameter. The first and second terms on the right hand side correspond to least-squares fitting errors for the EC and EO spectra, respectively, while the third is the regularization term to penalize differences between state-distinct parameters.

The amount of regularization applied (i.e. the value of λ) affects the quality of the fit. If regularization were too strong, it would force each (state-distinct) parameter to have the same value in EC and EO states, resulting in identical predicted spectra for each state and thus poor fit accuracy (assuming that the two spectra are actually different). If regularization were too weak, the parameter values vary too wildly, as we found in Fig B in S1 Appendix. Our strategy is to maximize the amount of regularization applied while keeping the fit inside the uncertainty bounds of the data.

To determine this optimal λ, we calculate fitting errors (the first two expressions on the right hand side of Eq 5) for 19 different values of λ ranging across ten orders of magnitude. The resulting plot (see Fig D in S1 Appendix) has an "S" shape, exhibiting high fitting accuracy at low λ and poor accuracy at high λ, with a transition regime in between. Our optimal regularization parameter is taken to be the largest value of λ where the median regularised fitting error does not exceed the 84% quantile of unregularized fitting errors. This corresponds to a value $\lambda = 0.1$. A visual comparison of regularized versus unregularized fits is given in Fig 2.

Supporting information

S1 Appendix. Additional figures. Fig A. Degree of alpha blocking across all subjects; Fig B. Unregularized EC to EO parameter responses and how they scale with the degree of alpha blocking; Fig C. Manhattan distances between EC and EO parameter sets as a function of the degree of alpha blocking; Fig D. Comparison of fitting error as a function of regularization parameter. (PDF)

References

1. Berger H. On the electroencephalogram of man. Third Report 1931; Twelfth Report 1937. Translated by Pierre Gloor. Electroencephalogr Clin Neurophysiol. 1931;28(suppl):113–167.

2. Adrian ED, Matthews BHC. The Berger Rhythm: Potential Changes from the Occipital Loves in Man. Brain. 1934;57(4):355–385. doi:10.1093/brain/57.4.355.

3. Chang BS, Schomer DL, Niedermeyer E. Normal EEG and sleep: adults and elderly. In: Schomer DL, da Silva FL, editors. Niedermeyer's

Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Philadelphia, PA: Lippincott Williams and Wilkins; 2010. p. 183–214.

4. Hanslmayr S, Gross J, Klimesch W, Shapiro KL. The role of alpha oscillations in temporal attention. Brain Research Reviews. 2011;67(1):331 – 343. doi:https://doi.org/10.1016/j.brainresrev.2011.04.002.

5. Başar E, Güntekin B. A short review of alpha activity in cognitive processes and in cognitive impairment. International Journal of Psychophysiology. 2012;86(1):25 – 38. doi:https://doi.org/10.1016/j.ijpsycho.2012.07.001.

6. Klimesch W. Alpha-band oscillations, attention, and controlled access to stored information. Trends in Cognitive Sciences. 2012;16(12):606 – 617. doi:https://doi.org/10.1016/j.tics.2012.10.007.

7. Lange J, Keil J, Schnitzler A, van Dijk H, Weisz N. The role of alpha oscillations for illusory perception. Behavioural Brain Research. 2014;271:294 – 301. doi:https://doi.org/10.1016/j.bbr.2014.06.015.

8. Lozano-Soldevilla D. On the Physiological Modulation and Potential Mechanisms Underlying Parieto-Occipital Alpha Oscillations. Frontiers in Computational Neuroscience. 2018;12:23. doi:10.3389/fncom.2018.00023.

9. Abeysuriya RG, Rennie CJ, Robinson PA. Physiologically based arousal state estimation and dynamics. J Neurosci Methods. 2015;253:55–69.

10. Vijayan S, Ching S, Purdon PL, Brown EN, Kopell NJ. Thalamocortical mechanisms for the anteriorization of alpha rhythms during propofol-induced unconsciousness. J Neurosci. 2013;33(27):11070–11075.

11. Jones SR, Pritchett DL, Sikora MA, Stufflebeam SM, Hamalainen M, Moore CI. Quantitative analysis and biophysically realistic neural modeling of the MEG mu rhythm: rhythmogenesis and modulation of sensory-evoked responses. J Neurophysiol. 2009;102(6):3554–3572.

12. Robinson PA, Rennie CJ, Rowe DL. Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. Phys Rev E. 2002;65:041924. doi:10.1103/PhysRevE.65.041924.

13. Nunez PL, Srinivasan R. Electric Fields of the Brain: The Neurophysics of EEG. 2nd ed. Oxford: Oxford University Press; 2005.

14. Elul R. The genesis of the EEG. Int Rev Neurobiol. 1972;15:227–272.

15. Moruzzi G. Reticular influences on the EEG. Electroencephalography and clinical neurophysiology. 1964;16:2–17.

16. Wilson HR, Cowan JD. Excitatory and Inhibitory Interactions in Localized Populations of Model Neurons. Biophysical Journal. 1972;12(1):1 – 24. doi:https://doi.org/10.1016/S0006-3495(72)86068-5.

17. Nunez PL, Cutillo BA. Neocortical Dynamics and Human EEG Rhythms. Oxford University Press; 1995. Available from: https://books.google.com.au/books?id=Xmq1AAAAIAAJ.

18. Deco G, Jirsa VK, Robinson PA, Breakspear M, Friston K. The Dynamic Brain: From Spiking Neurons to Neural Masses and Cortical Fields. PLOS Computational Biology. 2008;4(8):1–35. doi:10.1371/journal.pcbi.1000092.

19. Liley DT, Foster BL, Bojak I. Co-operative populations of neurons: mean field models of mesoscopic brain activity. In: Computational Systems Neurobiology. Springer; 2012. p. 317–364.

20. In: Jaeger D, Jung R, editors. Neural Population Models of the Alpha Rhythm. Encyclopedia of Computational Neuroscience. New York, NY: Springer New York; 2015. p. 1912–1912. Available from: https://doi.org/10.1007/978-1-4614-6675-8_100394.

21. Lopes da Silva FH, Hoeks A, Smits H, Zetterberg LH. Model of brain rhythmic activity: The alpha-rhythm of the thalamus. Kybernetik. 1974;15:27.

22. Liley DT, Cadusch PJ, Dafilis MP. A spatially continuous mean field theory of electrocortical activity. Network: Computation in Neural Systems. 2002;13(1):67–113.

23. Sotero RC, Trujillo-Barreto NJ, Iturria-Medina Y, Carbonell F, Jimenez JC. Realistically Coupled Neural Mass Models Can Generate EEG Rhythms. Neural Computation. 2007;19(2):478–512. doi:10.1162/neco.2007.19.2.478.

24. van Rotterdam A, Lopes da Silva FH, van den Ende J, Viergever MA, Hermans AJ. A model of the spatial-temporal characteristics of the alpha rhythm. Bulletin of Mathematical Biology. 1982;44:283–305. doi:https://doi.org/10.1007/BF02463252.

25. Robinson PA, Rennie CJ, Wright JJ. Propagation and stability of waves of electrical activity in the cerebral cortex. Phys Rev E. 1997;56:826–840. doi:10.1103/PhysRevE.56.826.

26. Liley DT, Cadusch PJ, Wright JJ. A continuum theory of electro-cortical activity. Neurocomputing. 1999;26:795–800.

27. Hughes SW, Crunelli V. Thalamic Mechanisms of EEG Alpha Rhythms and Their Pathological Implications. The Neuroscientist. 2005;11(4):357–372. doi:10.1177/1073858405277450.

28. Hughes SW, Crunelli V. Just a phase they're going through: The complex interaction of intrinsic high-threshold bursting and gap junctions in the generation of thalamic α and θ rhythms. International Journal of Psychophysiology. 2007;64(1):3 – 17. doi:https://doi.org/10.1016/j.ijpsycho.2006.08.004.

29. Freyer F, Roberts JA, Becker R, Robinson PA, Ritter P, Breakspear M. Biophysical Mechanisms of Multistability in Resting-State Cortical Rhythms. Journal of Neuroscience. 2011;31(17):6353–6361. doi:10.1523/JNEUROSCI.6693-10.2011.

30. Freyer F, Roberts JA, Ritter P, Breakspear M. A Canonical Model of Multistability and Scale-Invariance in Biological Systems. PLOS Computational Biology. 2012;8(8):1–15. doi:10.1371/journal.pcbi.1002634.

31. Rowe DL, Robinson PA, Rennie CJ. Estimation of neurophysiological parameters from the waking EEG using a biophysical model of brain dynamics. Journal of theoretical biology. 2004;231(3):413–433.

32. Hindriks R, van Putten MJAM. Thalamo-cortical mechanisms underlying changes in amplitude and frequency of human alpha oscillations. NeuroImage. 2013;70:150 – 163. doi:https://doi.org/10.1016/j.neuroimage.2012.12.018.

33. Hartoyo A, Cadusch PJ, Liley DT, Hicks DG. Parameter estimation and identifiability in a neural population model for electro-cortical activity. PLoS computational biology. 2019;15(5):e1006694.

34. Bellman R, Åström KJ. On structural identifiability. Mathematical biosciences. 1970;7(3-4):329–339.

35. Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics. 2014;30(10):1440–1448.

36. Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. Physical review E. 2003;68(2):021904.

37. Transtrum MK, Machta BB, Sethna JP. Why are nonlinear fits to data so challenging? Physical review letters. 2010;104(6):060201.

38. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215–e220.

39. Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Transactions on audio and electroacoustics. 1967;15(2):70–73.

40. Markand ON. Alpha rhythms. Journal of Clinical Neurophysiology. 1990;7(2):163–190.

41. Könönen M, Partanen J. Blocking of EEG alpha activity during visual performance in healthy adults. A quantitative study. Electroencephalography and clinical neurophysiology. 1993;87(3):164–166.

42. Bojak I, Liley D. Modeling the effects of anesthesia on the electroencephalogram. Physical Review E. 2005;71(4):041902.

43. He BJ, Zempel JM, Snyder AZ, Raichle ME. The temporal structures and functional significance of scale-free brain activity. Neuron. 2010;66(3):353–369.

44. Hastie T, Tibshirani R, Wainwright M. Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman & Hall/CRC; 2015.

45. Eberhart R, Kennedy J. A new optimizer using particle swarm theory. In: Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on. IEEE; 1995. p. 39–43.

46. Shi Y, Eberhart R. A modified particle swarm optimizer. In: Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on. IEEE; 1998. p. 69–73.

47. John ER, Prichep LS, Kox W, Valdés-Sosa P, Bosch-Bayard J, Aubert E, et al. Invariant Reversible QEEG Effects of Anesthetics. Consciousness and Cognition. 2001;10(2):165 – 183. doi:https://doi.org/10.1006/ccog.2001.0507.

48. Antkowiak B. In vitro networks: cortical mechanisms of anaesthetic action. British Journal of Anaesthesia. 2002;89(1):102 – 111. doi:https://doi.org/10.1093/bja/aef154.

49. Kullback S, Leibler RA. On information and sufficiency. The annals of mathematical statistics. 1951;22(1):79–86.

50. Kullback S. Information theory and statistics. Courier Corporation; 1997.

51. Lin J. Divergence measures based on the Shannon entropy. IEEE Transactions on Information theory. 1991;37(1):145–151.

52. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological). 1996;58(1):267–288.
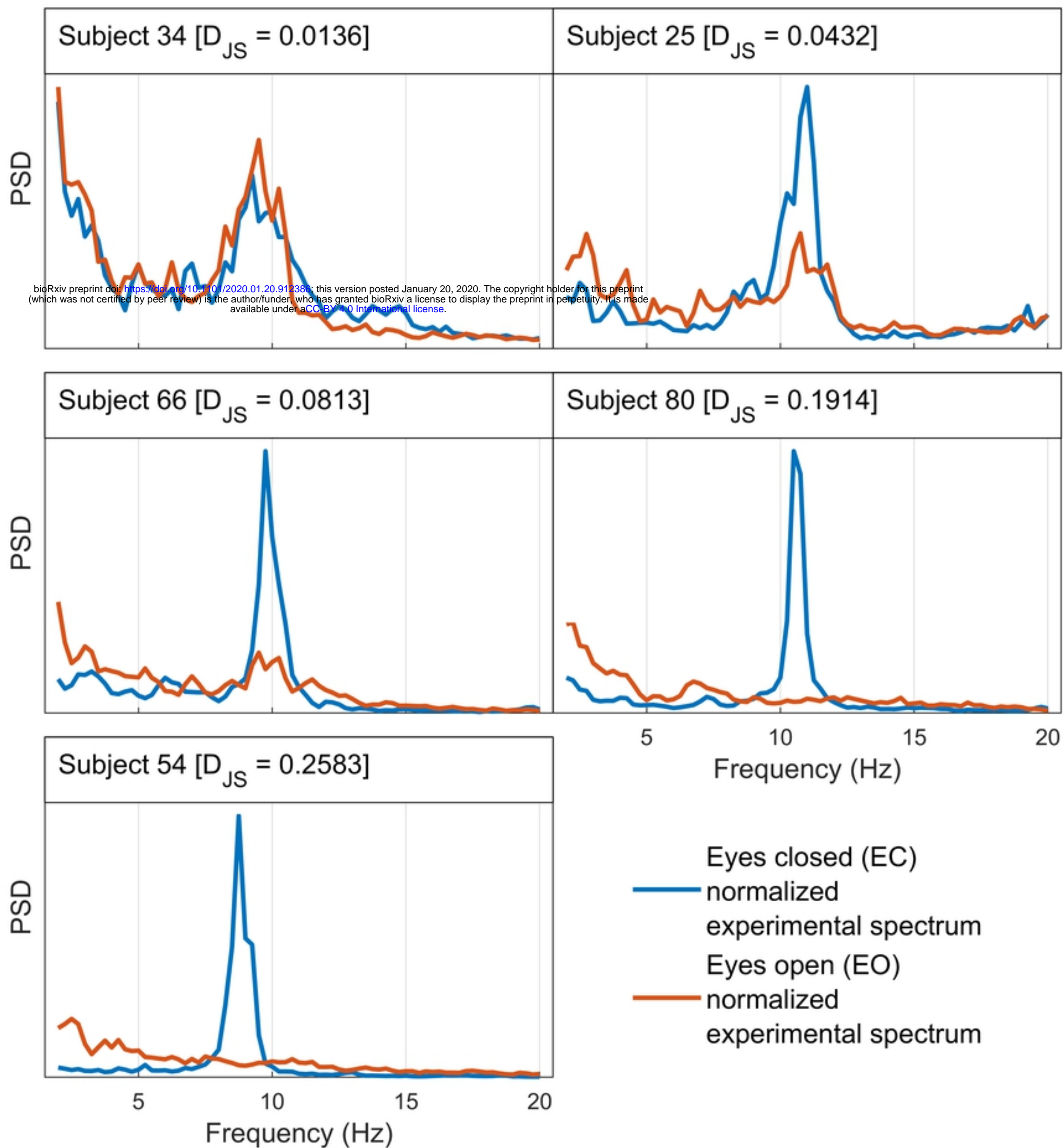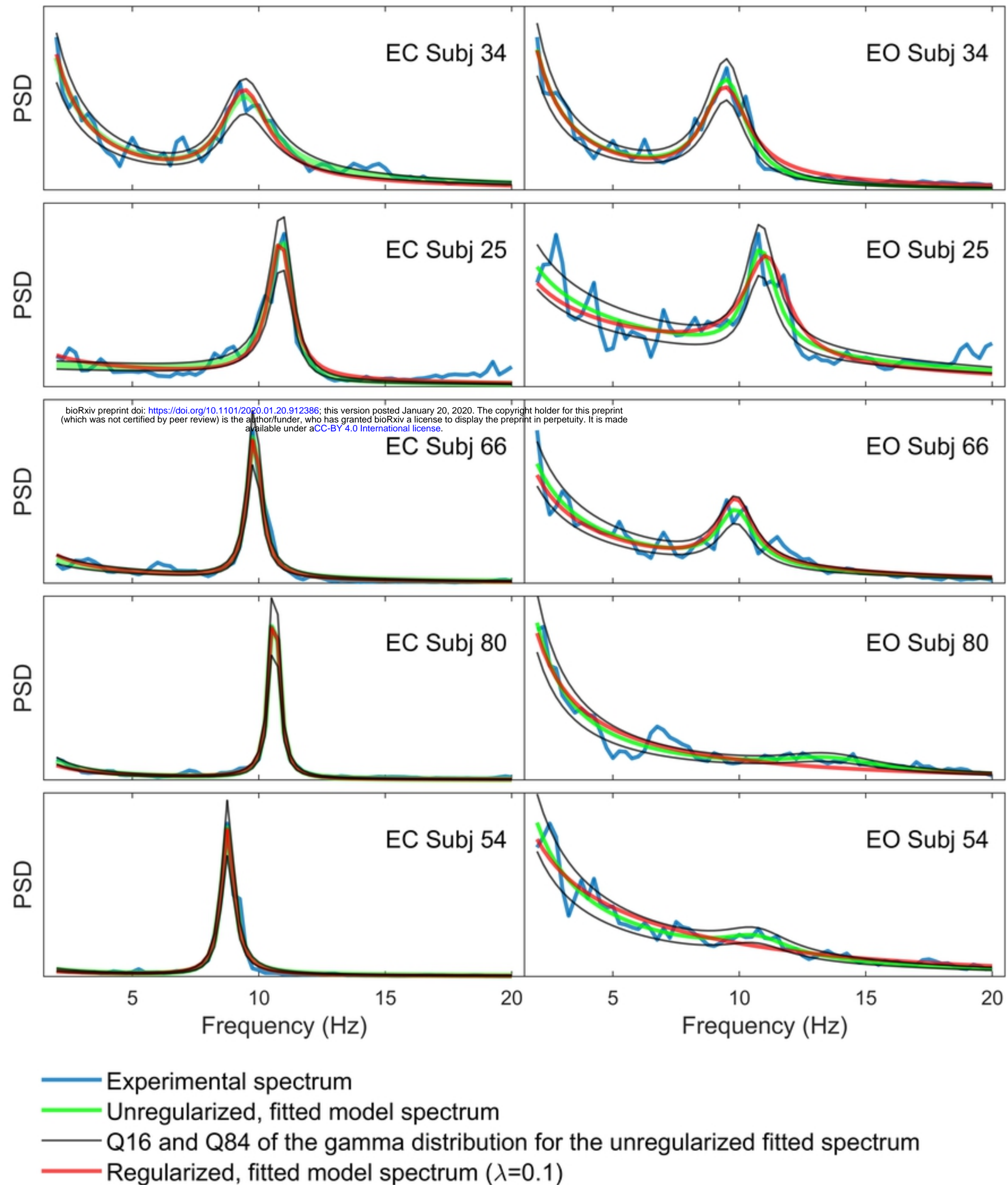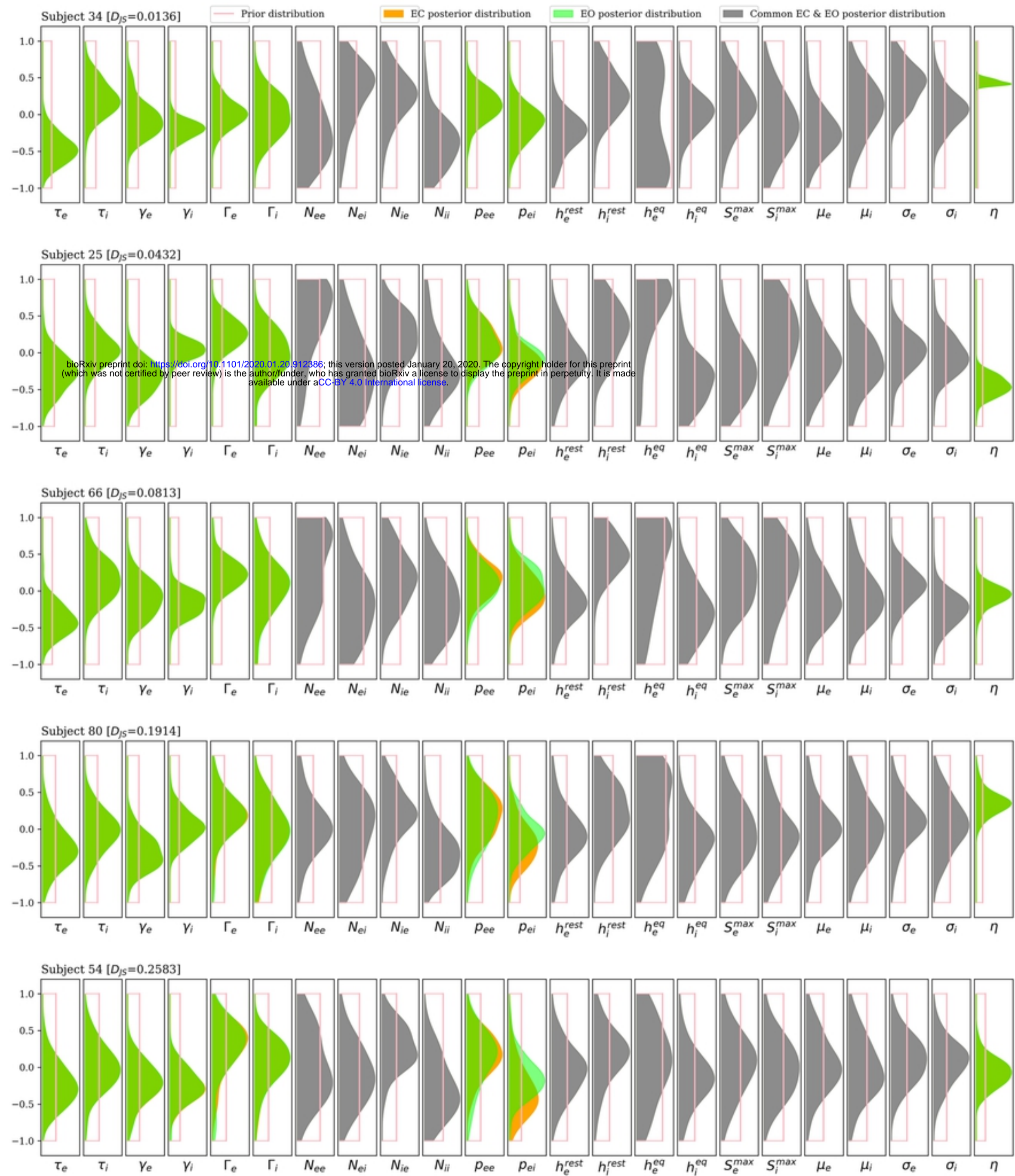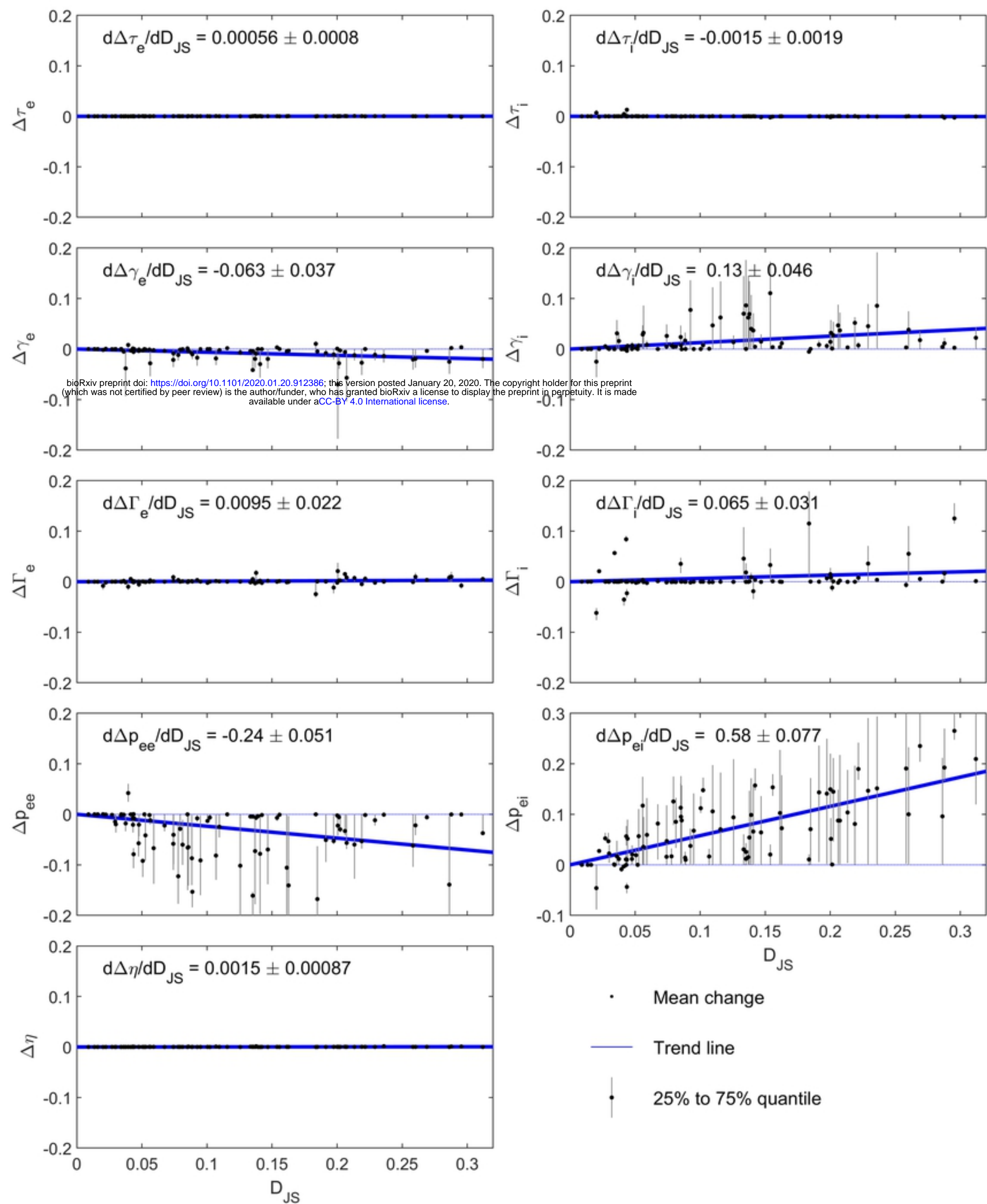
Fig 1

EC Subj 34 — EO Subj 34
EC Subj 25 — EO Subj 25
EC Subj 66 — EO Subj 66
EC Subj 80 — EO Subj 80
EC Subj 54 — EO Subj 54

Frequency (Hz)

PSD

— Experimental spectrum
— Unregularized, fitted model spectrum
— Q16 and Q84 of the gamma distribution for the unregularized fitted spectrum
— Regularized, fitted model spectrum (λ=0.1)

Fig 2

Fig 3

Fig 4

Experimental spectrum (EO Subject 25)

Fitted model spectrum: $\theta = \theta_0$

Perturbed model spectrum: $\theta = \theta_0 + 0.03(\theta_{max} - \theta_{min})$

Perturbed model spectrum: $\theta = \theta_0 - 0.03(\theta_{max} - \theta_{min})$

Fig 5