Information Geometry Theoretic Measures for Characterizing Neural Information Processing from Simulated EEG Signals

Hua, J-C., Kim, E. & He, F.

Article

Information Geometry Theoretic Measures for Characterizing Neural Information Processing from Simulated EEG Signals

Jia-Chen Hua [1,*] [ID] **, Eun-jin Kim** [1] [ID] **and Fei He** [2] [ID]

1 Centre for Fluid and Complex Systems, Coventry University, Coventry CV1 2NL, UK; ejk92122@gmail.com
2 Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 2TL, UK; fei.he@coventry.ac.uk
* Correspondence: jia-chen.hua@coventry.ac.uk

Abstract: In this work, we explore information geometry theoretic measures for characterizing neural information processing from EEG signals simulated by stochastic nonlinear coupled oscillator models for both healthy subjects and Alzheimer's disease (AD) patients with both eyes-closed and eyes-open conditions. In particular, we employ information rates to quantify the time evolution of probability density functions of simulated EEG signals, and employ causal information rates to quantify one signal's instantaneous influence on another signal's information rate. These two measures help us find significant and interesting distinctions between healthy subjects and AD patients when they open or close their eyes. These distinctions may be further related to differences in neural information processing activities of the corresponding brain regions, and to differences in connectivities among these brain regions. Our results show that information rate and causal information rate are superior to their more traditional or established information-theoretic counterparts, i.e., differential entropy and transfer entropy, respectively. Since these novel, information geometry theoretic measures can be applied to experimental EEG signals in a model-free manner, and they are capable of quantifying non-stationary time-varying effects, nonlinearity, and non-Gaussian stochasticity presented in real-world EEG signals, we believe that they can form an important and powerful tool-set for both understanding neural information processing in the brain and the diagnosis of neurological disorders, such as Alzheimer's disease as presented in this work.

Keywords: information geometry; information length; information rate; causal information rate; causality; stochastic oscillators; electroencephalography; stochastic simulation; signal processing; dementia; Alzheimer's disease; information theory; neural information processing; brain networks

1. Introduction

Identifying quantitative features from neurophysiological signals such as electroencephalography (EEG) is critical for understanding neural information processing in the brain and the diagnosis of neurological disorders such as dementia. Many such features have been proposed and employed to analyze neurological signals, which not only resulted in insightful understanding of the brain neurological dynamics of patients with certain neurological disorders versus healthy control (CTL) groups, but also helped build mathematical models that replicate the neurological signal with these quantitative features [1–5].

An important distinction, or non-stationary time-varying effects of the neurological dynamics, is the switching between eyes-open (EO) and eyes-closed (EC) states, where numerous research studies have been conducted on this distinction between EO and EC states to quantify important features of CTL subjects and patients using different techniques on EEG data such as traditional frequency-domain analysis [6,7], transfer entropy [8], energy landscape analysis [9], and nonlinear manifold learning for functional connectivity analysis [10], while also attempting to relate these features to specific clinical conditions and/or physiological variables, including skin conductance levels [11,12], cerebral blood

flow [13], brain network connectivity [14–16], brain activities in different regions [17], and performance on the unipedal stance test (UPST) [18]. Clinical physiological studies found that there are distinct mental states related to the EO and EC states. Specifically, there is an "exteroceptive" mental activity state characterized by attention and ocular motor activity during EO, and an "interoceptive" mental activity state characterized by imagination and multisensory activity during EC [19,20]. Ref. [21] suggested that the topological organization of human brain networks dynamically switches corresponding to the information processing modes when the brain is visually connected to or disconnected from the external environment. However, patients with Alzheimer's disease (AD) show loss of brain responsiveness to environmental stimuli [22,23], which might be due to impaired or loss of connectivities in the brain networks. This suggests that dynamical changes between EO and EC might represent an ideal paradigm to investigate the effect of AD pathophysiology and could be developed as biomarkers for diagnosis purposes. However, sensible quantification of robust features of these dynamical changes between EO and EC of both healthy and AD subjects, solely relying on EEG signals, is nontrivial. Despite the success of many statistical and quantitative measures being applied to neurological signal analysis, the main challenges stem from the non-stationary time-varying dynamics of the human brain with nonlinearity and non-Gaussian stochasticity, which makes most, if not all, of these traditional quantitative measures inadequate, and blindly applying these traditional measures to nonlinear and nonstationary time series/signals may produce spurious results, leading to incorrect interpretation.

In this work, by using simulated EEG signals of both CTL groups and AD patients under both EC and EO conditions and based on our previous works on information geometry [24–26], we develop novel and powerful quantitative measures in terms of information rate and causal information rate to quantify the important features of neurological dynamics of brains. We are able to find significant and interesting distinctions between CTL subjects and AD patients when they switch between the eyes-open and eyes-closed status. These quantified distinctions may be further related to differences in neural information processing activities of the corresponding brain regions, and to differences in connectivities among these brain regions, and therefore, they can be further developed as important biomarkers to diagnose neurological disorders, including but not limited to Alzheimer's disease. It should be noted that these novel and powerful quantitative measures in terms of information rate and causal information rate can be applied to experimental EEG signals in a model-free manner, and they are capable of quantifying non-stationary time-varying effects, nonlinearity, and non-Gaussian stochasticity presented in real-world EEG signals, and hence, they are more robust and reliable than other information-theoretic measures applied to neurological signal analysis in the literature [27,28]. Therefore, we believe that these information geometry theoretic measures can form an important and powerful tool set for the neuroscience community.

The EEG signals have been modeled using many different methodologies in the literature. An EEG model in terms of nonlinear stochastic differential equation (SDE) could be sufficiently flexible in that it usually contains many parameters, whose values can be tuned to match the model's output with actual EEG signals for different neurophysiological conditions, such as EC and EO, of CTL subjects or AD patients. Moreover, an SDE model of EEG can be solved by a number of numerical techniques to generate simulated EEG signals superior to actual EEG signals in terms of much higher temporal resolution and much larger number of sample paths available. These are the two main reasons why we choose to work with SDE models of EEG signals. Specifically, we employed a model of stochastic coupled Duffing–van der Pol oscillators proposed by Ref. [1], which is flexible enough to represent the EC and EO conditions for both CTL and AD subjects and straightforward enough to be simulated by using typical numerical techniques for solving SDE. Moreover, the model parameters reported in Ref. [1] were fine tuned against real-world experimental EEG signals of CTL and AD patients with both EC and EO conditions, and therefore,

quantitative investigations on the model's output of simulated signals are sufficiently representative for a large population of healthy and AD subjects.

2. Methods

2.1. Stochastic Nonlinear Oscillator Models of EEG Signals

A phenomenological model of the EEG based on a coupled system of Duffing–van der Pol oscillators subject to white noise excitation has been introduced [1] with the following form:

$$\begin{cases} \ddot{x}_1 + (k_1 + k_2)x_1 - k_2x_2 = -b_1x_1^3 - b_2(x_1 - x_2)^3 + \epsilon_1\dot{x}_1(1 - x_1^2), \\ \ddot{x}_2 - k_2x_1 + k_2x_2 = b_2(x_1 - x_2)^3 + \epsilon_2\dot{x}_2(1 - x_2^2) + \mu\,dW, \end{cases} \tag{1}$$

where $x_i, \dot{x}_i, \ddot{x}_i, i = 1, 2$ are positions, velocities, and accelerations of the two oscillators, respectively. Parameters $k_i, b_i, \epsilon_i, i = 1, 2$ are the linear stiffness, cubic stiffness, and van der Pol damping coefficient of the two oscillators, respectively. Parameter μ represents the intensity of white noise and dW is a Wiener process representing the additive noise in the stochastic differential system. The physical meanings of these variables and parameters were nicely explained in a schematic figure in Ref. [1].

By using actual EEG signals, Ref. [1] utilized a combination of several different statistical and optimization techniques to fine tune the parameters in the model equations for eyes-closed (EC) and eyes-open (EO) conditions of both healthy control (CTL) subjects and Alzheimer's disease (AD) patients, and these parameter values for different conditions are summarized in Tables 1 and 2.

Table 1. Optimal parameters of the Duffing–van der Pol oscillator for EC and EO of healthy control (CTL) subjects.

Parameter	Eyes-Closed (EC)	Eyes-Open (EO)
k_1	7286.5	2427.2
k_2	4523.5	499.92
b_1	232.05	95.61
b_2	10.78	103.36
ϵ_1	33.60	48.89
ϵ_2	0.97	28.75
μ	2.34	1.82

Table 2. Optimal parameters of the Duffing–van der Pol oscillator for EC and EO of Alzheimer's disease (AD) patients.

Parameter	Eyes-Closed (EC)	Eyes-Open (EO)
k_1	1742.1	3139.9
k_2	1270.8	650.32
b_1	771.99	101.1
b_2	1.91	81.3
ϵ_1	63.7	56.3
ϵ_2	20.7	19.12
μ	1.78	1.74

The model Equation (1) can be easily rewritten in a more standard form of stochastic differential equation (SDE) as follows:

$$\begin{cases} \dot{x}_1 = x_3, \\ \dot{x}_2 = x_4, \\ \dot{x}_3 = -(k_1 + k_2)x_1 + k_2 x_2 - b_1 x_1^3 - b_2(x_1 - x_2)^3 + \epsilon_1 x_3(1 - x_1^2), \\ \dot{x}_4 = k_2 x_1 - k_2 x_2 + b_2(x_1 - x_2)^3 + \epsilon_2 x_4(1 - x_2^2) + \mu \, dW, \end{cases} \tag{2}$$

which is more readily suitable for stochastic simulations.

2.2. Initial Conditions (ICs) and Specifications of Stochastic Simulations

For simplicity, we employ the Euler–Maruyama scheme [29] to simulate 2×10^7 trajectories in total of the model Equation (2); although, other more sophisticated methods for stochastic simulations exist. We simulate such a large number of trajectories, because calculations of information geometry theoretic measures rely on accurate estimation of probability density functions (PDFs) of the model's variables $x_i(t)$, which requires a large number of data samples of $x_i(t)$ at any given time t.

Since nonlinear oscillators' solution is very sensitive to initial conditions, we start the simulation with a certain initial probability distribution (e.g., a Gaussian distribution) for all $x_1(0), x_2(0), x_3(0), x_4(0)$, which means that the 20 million $x_i(0)$ ($\forall i = 1, 2, 3, 4$) are randomly drawn from a probability density function (PDF) of the initial distribution. The time-step size dt is set to 10^{-6} to compensate for the very-high values of stiffness parameters k_1 and k_2 in Tables 1 and 2. The total number of simulation time steps is 1×10^7, making the total time range of simulation $[0, 10]$. The $\Delta t = 10^{-4}$ is the time interval when the probability density functions (PDFs) $p(x_1, t)$ and $p(x_2, t)$ are estimated for calculating information geometry theoretic measures such as information rates and causal information rates, as explained in Section 2.3.

For nonlinear oscillators, different initial conditions can result in dramatically different long-term time evolution. So in order to explore more diverse initial conditions, we simulated the SDE with 6 different initial Gaussian distributions with different means and standard deviations, i.e., $x_1(0) \sim \mathcal{N}(\mu_{x_1(0)}, \sigma^2)$, $x_2(0) \sim \mathcal{N}(\mu_{x_2(0)}, \sigma^2)$, $x_3(0) \sim \mathcal{N}(\mu_{x_3(0)}, \sigma^2)$, $x_4(0) \sim \mathcal{N}(\mu_{x_4(0)}, \sigma^2)$, where the parameters are summarized alongside other specifications in Table 3.

Table 3. Initial conditions (IC): $x_1(0), x_2(0), x_3(0), x_4(0)$ are randomly drawn from Gaussian distributions $\mathcal{N}(\mu_{x_i(0)}, \sigma^2)$ with different $\mu_{x_i(0)}$'s and σ's ($i = 1, 2, 3, 4$).

	IC No.1	IC No.2	IC No.3	IC No.4	IC No.5	IC No.6
$\mu_{x_1(0)}$	1.0	0.9	0.2	0.1	0.5	0.2
$\mu_{x_2(0)}$	0.5	0.1	0.5	0.5	0.9	0.9
$\mu_{x_3(0)}$	0	1.0	0.5	0.2	1.0	0.1
$\mu_{x_4(0)}$	0	0.5	1.0	1.0	0.8	0.5
σ	0.1	0.1	0.1	0.5	0.5	0.5
Num. of trajectories	2×10^7	2×10^7	2×10^7	2×10^7	2×10^7	2×10^7
dt	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}	10^{-6}
Δt	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
Num. of time-steps	1×10^7	1×10^7	1×10^7	1×10^7	1×10^7	1×10^7
Total range of t	$[0, 10]$	$[0, 10]$	$[0, 10]$	$[0, 10]$	$[0, 10]$	$[0, 10]$

For brevity, in this paper, we use the word "initial conditions" or its abbreviation "IC" to refer to the (set of 4) initial Gaussian distributions from which the 20 million $x_i(0)$

($\forall i = 1, 2, 3, 4$) are randomly drawn. For example, the "IC No.6" in Table 3 (and simply "IC6" elsewhere in this paper) refers to the 6th (set of 4) Gaussian distributions with which we start the simulation, and the specifications of this stimulation are listed in the last column of Table 3.

2.3. Information Geometry Theoretic Measures: Information Rate and Causal Information Rate

When a stochastic differential equation (SDE) model exhibits non-stationary time-varying effects, nonlinearity, and/or non-Gaussian stochasticity, while we are interested in large fluctuations and extreme events in the solutions, simple statistics such as mean and variance might not suffice to compare the solutions of different SDE models (or same model with different parameters). In such cases, quantifying and comparing the time evolution of probability density functions (PDFs) of solutions will provide us with more information [30]. The time evolution of PDFs can be studied and compared through the framework of information geometry [31], wherein PDFs are considered as points on a Riemannian manifold (which is called the statistical manifold), and their time evolution can be considered as a motion on this manifold. Several different metrics can be defined on a probability space to equip it with a manifold structure, including a metric related to the Fisher Information [32], known as the Fisher Information metric [33,34], which we use in this work:

$$g_{\mu\nu}(\theta) \overset{\text{def}}{=} \int_X \frac{\partial \log p(x; \{\theta\})}{\partial \theta_\mu} \frac{\partial \log p(x; \{\theta\})}{\partial \theta_\nu} p(x; \{\theta\}) \, dx. \tag{3}$$

Here, $p(x; \{\theta\})$ denotes a continuous family of PDFs parameterized by parameters $\{\theta\}$. If a time-dependent PDF $p(x, t)$ is considered as a continuous family of PDFs parameterized by a single parameter time t, the metric tensor is then reduced to a scalar metric:

$$g(t) = \int dx \, \frac{1}{p(x, t)} \left[\frac{\partial p(x, t)}{\partial t} \right]^2. \tag{4}$$

The infinitesimal distance $d\mathcal{L}$ on the manifold is then given by $d\mathcal{L}^2 = g(t) \, dt^2$, where $\mathcal{L}$ is called the Information Length and defined as follows:

$$\mathcal{L}(t) \overset{\text{def}}{=} \int_0^t dt_1 \sqrt{\int dx \, \frac{1}{p(x, t_1)} \left[\frac{\partial p(x, t_1)}{\partial t_1} \right]^2}. \tag{5}$$

The Information Length $\mathcal{L}$ represents the dimensionless distance, which measures the total distance traveled on the statistical manifold. The time derivative of $\mathcal{L}$ then represents the speed of motion on this manifold:

$$\Gamma(t) \overset{\text{def}}{=} \lim_{dt \to 0} \frac{d\mathcal{L}(t)}{dt} = \sqrt{\int dx \, \frac{1}{p(x, t)} \left[\frac{\partial p(x, t)}{\partial t} \right]^2}, \tag{6}$$

which is referred to as the Information Rate. If multiple variables are involved, such as $x_i(t)$ where $i = 1, 2, 3, 4$ as in the stochastic nonlinear oscillator model Equation (2), we will use subscript in $\Gamma(t)$, e.g., $\Gamma_{x_2}(t)$ to denote the information rate of signal $x_2(t)$.

The notion of Causal Information Rate was introduced in Ref. [25] to quantify how one signal instantaneously influences another signal's information rate. As an example, the causal information rate of signal $x_1(t)$ influencing signal $x_2(t)$'s information rate is denoted and defined by $\Gamma_{x_1 \to x_2}(t) \overset{\text{def}}{=} \Gamma_{x_2}^*(t) - \Gamma_{x_2}(t)$, where

$$\Gamma_{x_2}(t)^2 = \int dx_2 \, p(x_2, t) [\partial_t \ln p(x_2, t)]^2, \tag{7}$$

and

$$\Gamma_{x_2}^*(t)^2 \overset{\text{def}}{=} \lim_{t_* \to t^+} \int dx_1 \, dx_2 \, p(x_2, t_*; x_1, t)[\partial_{t_*} \ln p(x_2, t_* | x_1, t)]^2$$

$$= \lim_{t_* \to t^+} \int dx_1 \, dx_2 \, p(x_2, t_*; x_1, t)[\partial_{t_*} \ln p(x_2, t_*; x_1, t)]^2, \qquad (8)$$

where the relation between conditional, joint, and marginal PDFs $p(x_2, t_* | x_1, t) = \frac{p(x_2, t_*; x_1, t)}{p(x_1, t)}$ and the fact $\partial_{t_*} p(x_1, t) = 0$ for $t_* \neq t$ are used in the 2nd equal sign above. $\Gamma_{x_2}^*$ denotes the (auto) contribution to the information rate from x_2 itself, while x_1 is given/known and frozen in time. In other words, $\Gamma_{x_2}^*$ represents the information rate of x_2 when the additional information of x_1 (at the same time with x_2) becomes available or known. Subtracting Γ_{x_2} from $\Gamma_{x_2}^*$ following the definition of $\Gamma_{x_1 \to x_2}$ then gives us the contribution of (knowing the additional information of) x_1 to Γ_{x_2}, signifying how x_1 instantaneously influences the information rate of x_2. One can easily verify that if signals $x_1(t)$ and $x_2(t)$ are statistically independent such that the equal-time joint PDF can be separated as $p(x_1, t; x_2, t) = p(x_1, t) \cdot p(x_2, t)$, then $\Gamma_{x_2}^*(t)$ will reduce to $\Gamma_{x_2}(t)$, making the causal information rate $\Gamma_{x_1 \to x_2} = 0$, which is consistent with the assumption that $x_1(t)$ and $x_2(t)$ are statistically independent at the same time t.

For numerical estimation purposes, one can derive simplified equations $\Gamma_{x_2}(t)^2 = 4 \int dx_2 \left[\partial_t \sqrt{p(x_2, t)} \right]^2$ and $\Gamma_{x_2}^*(t)^2 = 4 \lim_{t_* \to t^+} \int dx_1 \, dx_2 \left[\partial_{t_*} \sqrt{p(x_2, t_*; x_1, t)} \right]^2$ to ease the numerical calculations and avoid numerical errors in PDFs (due to finite sample-size estimations using a histogram-based approach) being doubled or enlarged when approximating the integrals in the original Equations (7) and (8) by finite summation. On the other hand, the time derivatives of the square root of PDFs are approximated by using temporally adjacent PDFs with each pair of two adjacent PDFs being separated by $\Delta t = 10^{-4}$ in time, as mentioned at the end of Section 2.2.

2.4. Shannon Differential Entropy and Transfer Entropy

As a comparison with more traditional and established information-theoretic measures, we also calculate differential entropy and transfer entropy using the numerically estimated PDFs and compare them with information rate and causal information rate, respectively.

The Shannon differential entropy of a signal $x(t)$ is defined to extend the idea of Shannon discrete entropy as

$$h(x(t)) = E[-\ln p(x, t)] = -\int p(x, t) \ln p(x, t) \, dx = -\int P(dx(t)) \ln \frac{P(dx(t))}{\mu(dx)}, \qquad (9)$$

where $\mu(dx) = dx$ is the Lebesgue measure, and $P(dx(t)) = p(x, t) \, \mu(dx) = p(x, t) \, dx$ is the probability measure. In other words, differential entropy is the negative relative entropy (Kullback-Leibler divergence) from the Lebesgue measure (considered as an unnormalized probability measure) to a probability measure P (with density p). In contrast, information rate $\Gamma_x(t) = \sqrt{\int dx \, p(x, t)[\partial_t \ln p(x, t)]^2} = \sqrt{\lim_{dt \to 0} \frac{2}{dt^2} \int dx \, p(x, t + dt) \ln \frac{p(x, t + dt)}{p(x, t)}}$ (see Ref. [24–26] for detailed derivations) is related to the rate of change in relative entropy of two infinitesimally close PDFs $p(x, t)$ and $p(x, t + dt)$. Therefore, although differential entropy can measure the complexity of a signal $x(t)$ at time t, it neglects how the signal's PDF $p(x, t)$ changes instantaneously at that time, which is crucial to quantify how new information can be reflected from the instantaneous entropy production rate of the signal $x(t)$. This is the theoretical reason why the information rate is a much better and more appropriate measure than differential entropy for characterizing the neural information processing from EEG signals of the brain, and the practical reason for this will be illustrated in terms of numerical results and discussed at the end of Sections 3.3.2 and 3.3.3.

The transfer entropy (TE) measures the directional flow or transfer of information between two (discrete-time) stochastic processes. The transfer entropy from a signal $x_1(t)$ to another signal $x_2(t)$ is the amount of uncertainty reduced in future values of $x_2(t)$ by knowing the past values of $x_1(t)$ given past values of $x_2(t)$. Specifically, if the amount of information is measured using Shannon's (discrete) entropy $H(X_t) = -\sum_x p(x,t) \log_2 p(x,t)$ of a stochastic process X_t and conditional entropy $H(Y_{t_2}|X_{t_1}) = -\sum_{x,y} p(x,t_1;y,t_2) \log_2 p(y,t_2|x,t_1)$, the transfer entropy from a process X_t to another process Y_t (for discrete-time $t \in \mathbb{Z}$) can be written as follows:

$$
\begin{aligned}
\mathrm{TE}_{X_t \to Y_t}(t) &= H\Big(Y_{t+1} \mid Y_{t:t-(k-1)}\Big) - H\Big(Y_{t+1} \mid Y_{t:t-(k-1)}, X_{t:t-(l-1)}\Big), \tag{10} \\
&= -\sum_y p(y_{t+1}, y_{t:t-(k-1)}) \log_2 p(y_{t+1}|y_{t:t-(k-1)}) \\
&\quad + \sum_{x,y} p(y_{t+1}, y_{t:t-(k-1)}, x_{t:t-(l-1)}) \log_2 p(y_{t+1}|y_{t:t-(k-1)}, x_{t:t-(l-1)}), \\
&= \sum_{x,y} p(y_{t+1}, y_{t:t-(k-1)}, x_{t:t-(l-1)}) \log_2 \frac{p(y_{t+1}|y_{t:t-(k-1)}, x_{t:t-(l-1)})}{p(y_{t+1}|y_{t:t-(k-1)})}, \\
&= \sum_{x,y} p(y_{t+1}, y_{t:t-(k-1)}, x_{t:t-(l-1)}) \log_2 \frac{p(y_{t+1}, y_{t:t-(k-1)}, x_{t:t-(l-1)}) p(y_{t:t-(k-1)})}{p(y_{t:t-(k-1)}, x_{t:t-(l-1)}) p(y_{t+1}, y_{t:t-(k-1)})}, \tag{11}
\end{aligned}
$$

which quantifies the amount of reduced uncertainty in future value Y_{t+1} by knowing the past l values of X_t given past k values of Y_t, where $Y_{t:t-(k-1)}$ and $X_{t:t-(l-1)}$ are shorthands of past k values $Y_t, Y_{t-1}, \ldots, Y_{t-(k-1)}$ and past l values $X_t, X_{t-1}, \ldots, X_{t-(l-1)}$, respectively.

In order to properly compare with causal information rate signifying how one signal instantaneously influences another signal's information rate (at the same/equal-time t), we set $k = l = 1$ in calculating the transfer entropy between two signals. Also, since the causal information rate involves partial time derivatives, which have to be numerically estimated using temporally adjacent PDFs separated by $\Delta t = 10^{-4}$ in time (as mentioned at the end of Section 2.2), the discrete-time $t \in \mathbb{Z}$ in transfer entropy should be changed to $n\Delta t$ with $n \in \mathbb{Z}$. Therefore, the transfer entropy appropriate for comparing with the causal information rate should be rewritten as follows:

$$
\begin{aligned}
\mathrm{TE}_{x_1 \to x_2}(t) &= H(x_2(t + \Delta t) \mid x_2(t)) - H(x_2(t + \Delta t) \mid x_2(t), x_1(t)), \tag{12} \\
&= \sum_{x_1, x_2} p(x_2, t + \Delta t; x_2, t; x_1, t) \log_2 \frac{p(x_2, t + \Delta t|x_2, t; x_1, t)}{p(x_2, t + \Delta t|x_2, t)}, \\
&= \sum_{x_1, x_2} p(x_2, t + \Delta t; x_2, t; x_1, t) \log_2 \frac{p(x_2, t + \Delta t; x_2, t; x_1, t) p(x_2, t)}{p(x_2, t; x_1, t) p(x_2, t + \Delta t; x_2, t)}. \tag{13}
\end{aligned}
$$

Numerical estimations of the information rate, causal information rate, differential entropy, and transfer entropy are all based on numerical estimation of PDFs using histograms. In particular, in order to sensibly and consistently estimate the causal information rate (e.g., to avoid getting negative values), special caution is required when choosing the binning for histogram estimation of PDFs in calculating $\Gamma_{x_2}(t)^2 = 4 \int dx_2 \left[\partial_t \sqrt{p(x_2, t)} \right]^2$ and $\Gamma_{x_2}^*(t)^2 = 4 \lim_{t_* \to t^+} \int dx_1 \, dx_2 \left[\partial_{t_*} \sqrt{p(x_2, t_*; x_1, t)} \right]^2$. The finer details for these numerical estimation techniques are elaborated in Appendix A.

3. Results

We performed simulations with six different Gaussian initial distributions (with different means and standard deviations summarized in Table 3). Initial Conditions No.1 (IC No.1, or simply IC1) through No.3 (IC3) are Gaussian distributions with a narrow width or smaller standard deviation, whereas IC4 through IC6 have a larger width/standard deviation, and therefore, the simulation results of IC4 through IC6 exhibit more diverse

time evolution behaviors (e.g., more complex attractors, as explained next), and hence, the corresponding calculation results are more robust or insensitive to the specific mean values $\mu_{x_i(0)}$'s of the initial Gaussian distributions (see Table 3 for more details). Therefore, in the main text here, we focus on these results from initial Gaussian distributions with wider width/larger standard deviation, and we list complete results from all six initial Gaussian distributions in the Appendix B. Specifically, we found that the results from IC4 through IC6 are qualitatively the same or very similar, and therefore, in the main text here, we illustrate and discuss the results from Initial Conditions No.4 (IC4), which is sufficiently representative for IC5 and IC6, and refer to other IC's (by referencing the relevant sections in Appendix B or explicitly illustrating the results) if needed.

3.1. Sample Trajectories of $X_1(T)$ and $X_2(T)$

To give a basic idea of how the simulated trajectories evolve in time, we start by illustrating 50 sample trajectories of $x_1(t)$ and $x_2(t)$ from the total 2×10^7 simulated trajectories for both CTL subjects and AD patients with both EC and EO conditions, which are visualized in Figures 1 and 2. Notice that from Figure 1c, one can see that it takes some time for the trajectories of $x_2(t)$ to settle down on some complex attractors for EC, which suggests a longer memory associated with EC of CTL. This is more evident as shown in the time evolution of PDF $p(x_2, t)$ in Figure 3c below.

(a) CTL EC $x_1(t)$
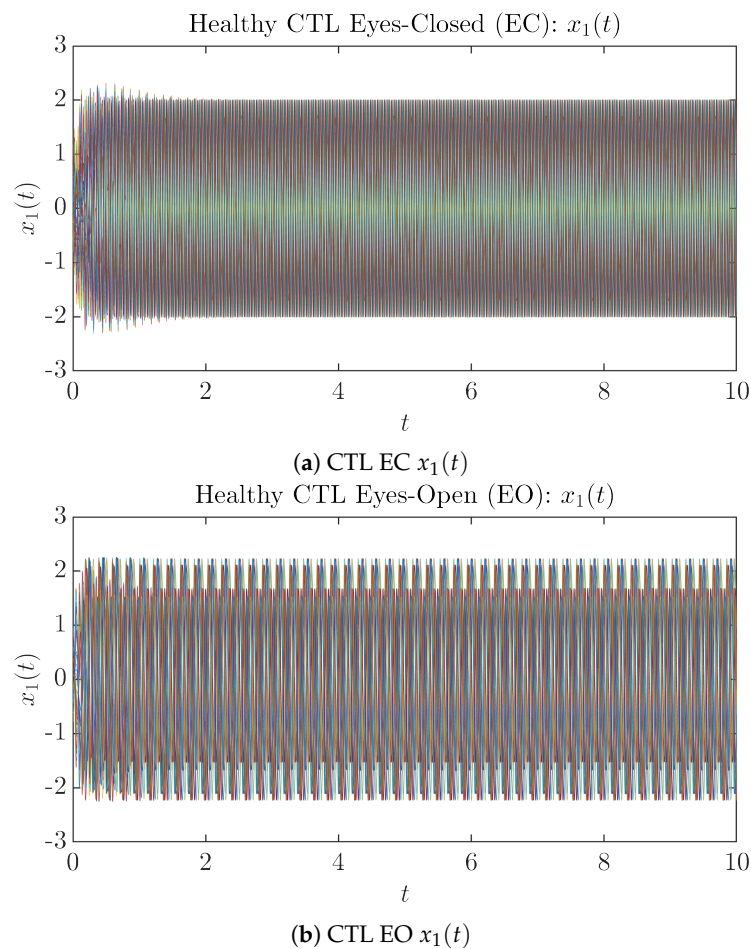
(b) CTL EO $x_1(t)$
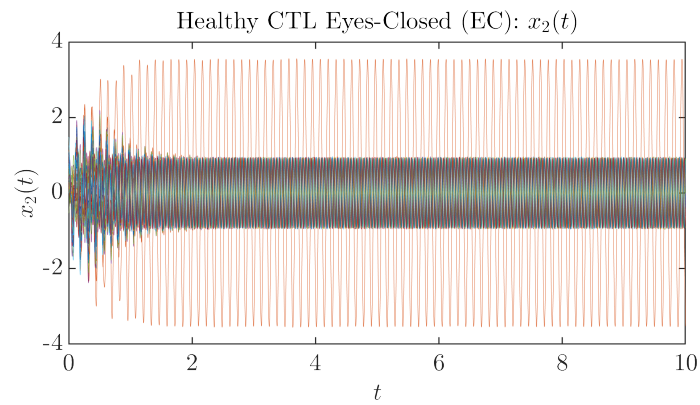
Figure 1. *Cont.*

(c) CTL EC $x_2(t)$
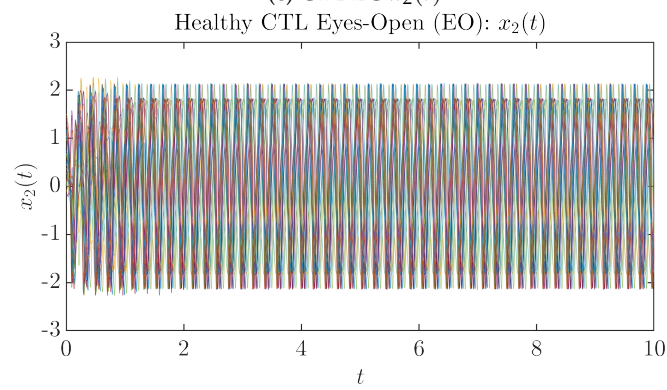
(d) CTL EO $x_2(t)$

Figure 1. Fifty Sample trajectories of healthy CTL subjects. Each single trajectory is labeled by a different color.
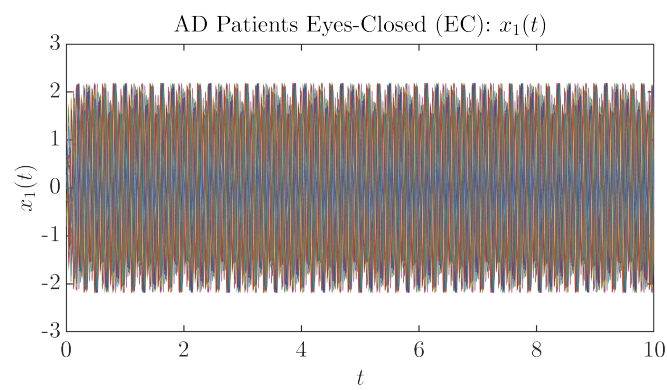
(a) AD EC $x_1(t)$
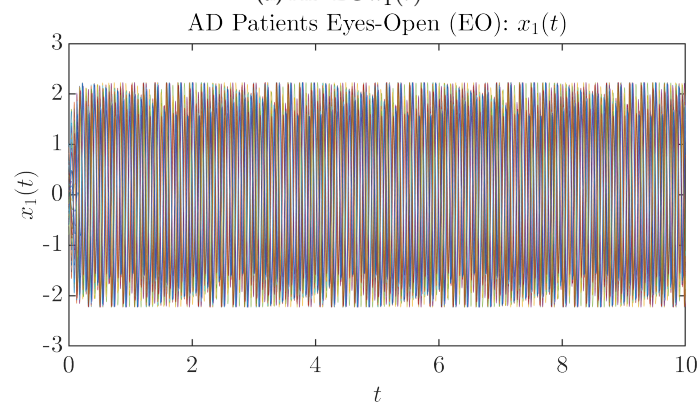
(b) AD EO $x_1(t)$

Figure 2. *Cont.*

(c) AD EC $x_2(t)$

(d) AD EO $x_2(t)$
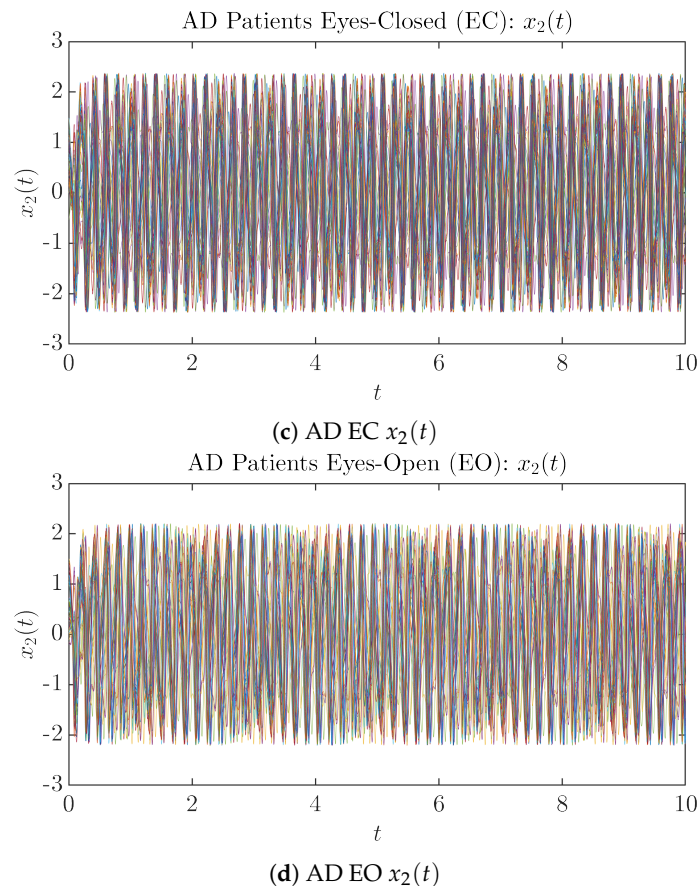
Figure 2. Fifty Sample trajectories of AD patients. Each single trajectory is labeled by a different color.

3.2. Time Evolution of PDF $P(X_1, T)$ and $P(X_2, T)$

The empirical PDFs $p(x_1, t)$ and $p(x_2, t)$ can better illustrate the overall time evolution of a large number of trajectories, and they serve as a basis for calculations of information geometry theoretic measures such as information rates and causal information rates. These empirical PDFs are estimated using a histogram-based approach with Rice's rule [35,36], where the number of bins is $n_{\text{bins}} = 2\sqrt[3]{n_{\text{samples}}}$, and since we simulated 2×10^7 sample trajectories in total, the n_{bins} is rounded to 542. The centers of bins are plotted on the y-axis in sub-figures of Figures 3 and 4, where the function values of $p(x_1, t)$ and $p(x_2, t)$ are color-coded following the color bars.

As mentioned in the previous section, from Figure 3c, one can see more clearly that after around $t \geq 5$, the trajectories settle down on some complex attractors, and the time evolution of $p(x_2, t)$ undergoes only minor changes. Meanwhile, from Figure 3a, one can observe that a similar settling down of $x_1(t)$ on some complex attractors happens after around $t \geq 7.5$. Therefore, we select only PDFs with $t \geq 7.5$ for statistical analysis of information rates and causal information rates to investigate the stationary properties.

From Figures 3 and 4, one can already observe some qualitative differences between healthy control (CTL) subjects and AD patients. For example, the time evolution patterns of $p(x_1, t)$ and $p(x_2, t)$ are significantly different when CTL subjects open their eyes from eyes-closed (EC) state, whereas for AD patients, these differences are relatively minor. One of the best ways to provide quantitative descriptions of these differences (instead of being limited to qualitative descriptions) is using information geometry theoretic measures such as information rates and causal information rates, whose results are listed in Sections 3.3 and 3.4, respectively.
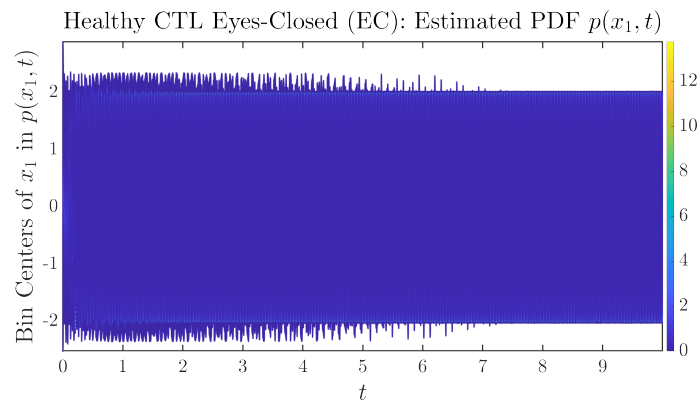
(a) CTL EC $p(x_1, t)$
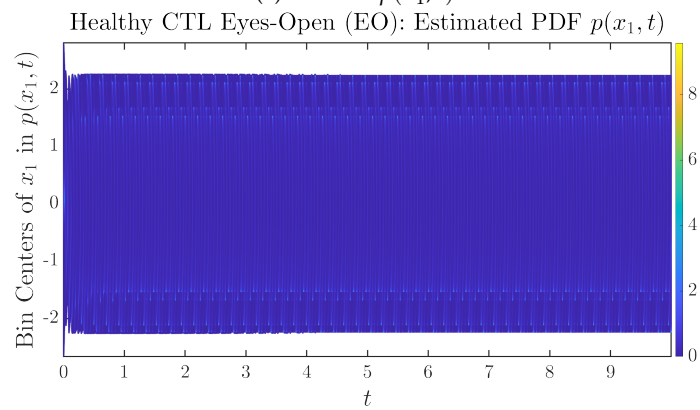
(b) CTL EO $p(x_1, t)$
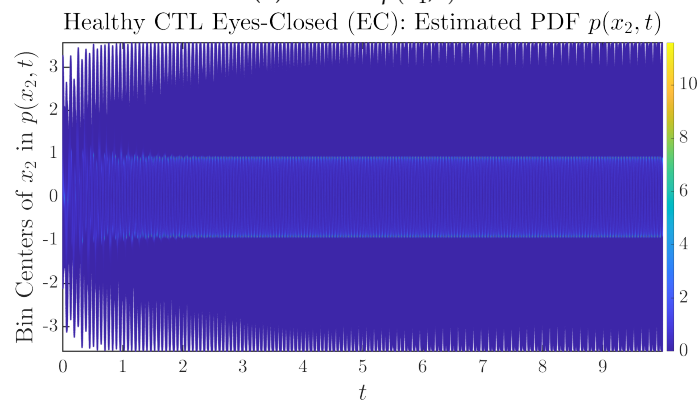
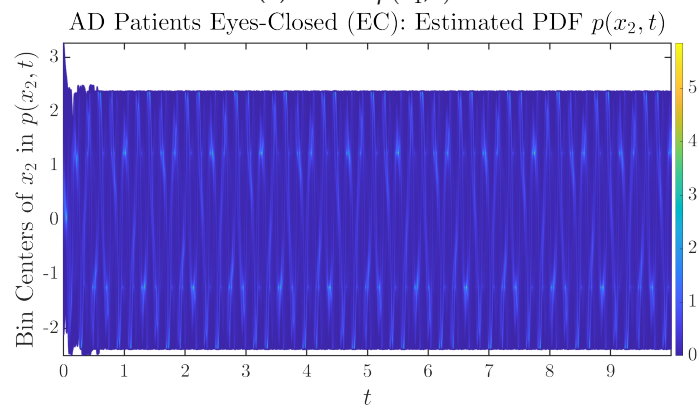**(c)** CTL EC $p(x_2, t)$
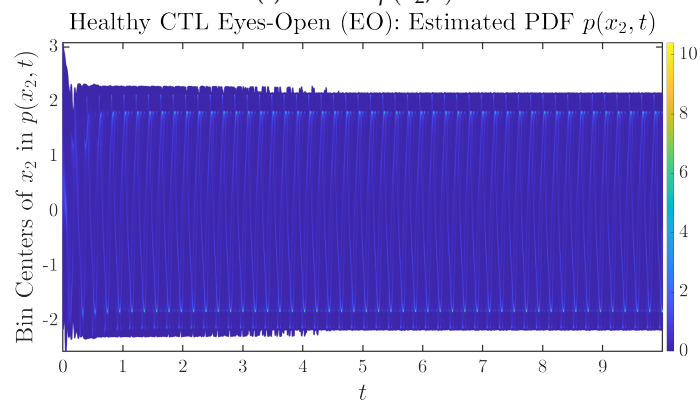
(d) CTL EO $p(x_2, t)$

Figure 3. Time evolution of estimated PDFs of healthy CTL subjects.
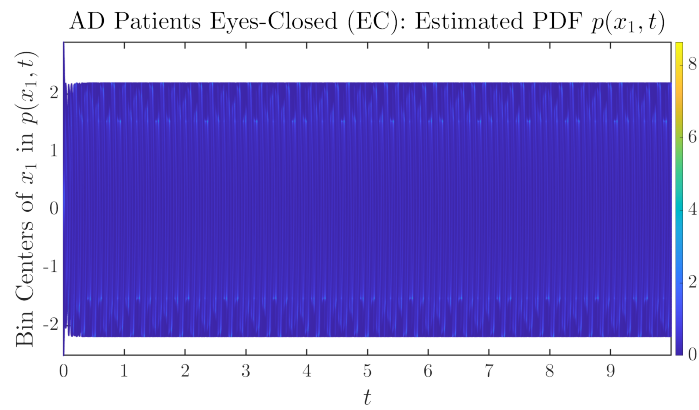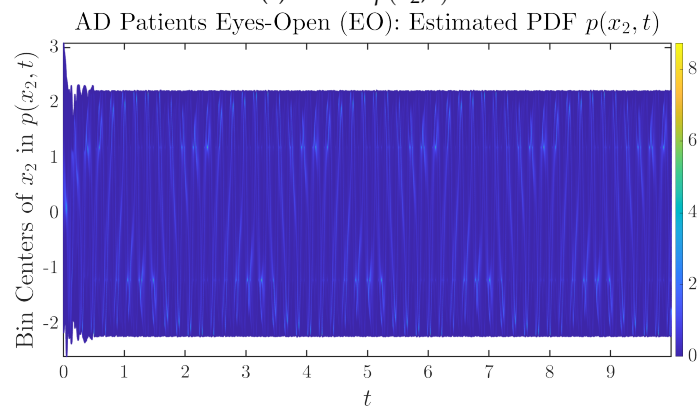
(a) AD EC $p(x_1, t)$
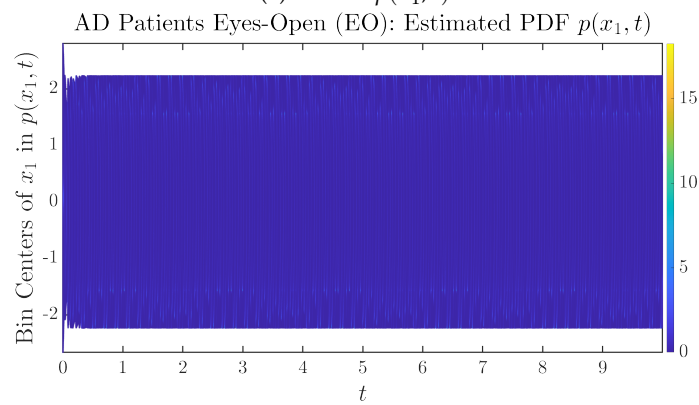
(b) AD EO $p(x_1, t)$

(c) AD EC $p(x_2, t)$

(d) AD EO $p(x_2, t)$

Figure 4. Time evolution of estimated PDFs of AD patients.

As can be seen from Appendix B.2, IC1 through IC3 exhibit much simpler attractors than IC4 through IC6. Since the width/standard deviation of the initial Gaussian distributions of IC1 through IC3 is much smaller, they are more sensitive to the specific mean values $\mu_{x_i(0)}$'s of the initial Gaussian distribution, and one can see that IC3's time evolution behaviors of $p(x_i, t)$ are somewhat qualitatively different from IC1 and IC2, whereas $p(x_i, t)$'s time evolution behaviors of IC4 through IC6 are all qualitatively the same.

3.3. Information Rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$

Intuitively speaking, the information rate is the (instantaneous) speed of PDF's motion on the statistical manifold, as each given PDF corresponds to a point on that manifold, and when the time changes, a time-dependent PDF will typically move on a curve on the statistical manifold, whereas a stationary or equilibrium state PDF will remain at the same point on the manifold. Therefore, the information rate is a natural tool to investigate the time evolution of PDF.

Moreover, since the information rate is quantifying instantaneous rate of change in the infinitesimal relative entropy between two adjacent PDFs, it is hypothetically a reflection of neural information processing in the brain, and hence, it may provide important insight into the neural activities in different regions of the brain, as long as the regional EEG signals can be sufficiently collected for calculating the information rates.

3.3.1. Time Evolution

The time evolution of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ are shown in Figure 5a,b for CTL subjects and AD patients, respectively. Since $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ quantify the (infinitesimal) relative entropy production rate instantaneously at time t, they represent the information-theoretic complexities of signals $x_1(t)$ and $x_2(t)$ of the coupled oscillators, respectively, and are hypothetical reflections of neural information processing in the corresponding regions in the brain.

For example, in Figure 5a, there is a clear distinction between eyes-closed (EC) and eyes-open (EO) for CTL subjects: both $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ decrease significantly when healthy subjects open their eyes, which may be interpreted as the neural information processing activities of the corresponding brain regions being "suppressed" by the incoming visual information when eyes are opened from being closed.
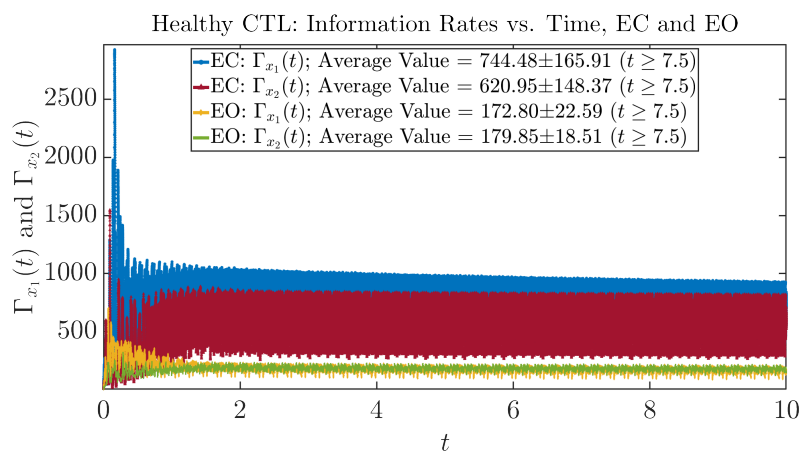
(a) Health CTL subjects.

Figure 5. *Cont.*
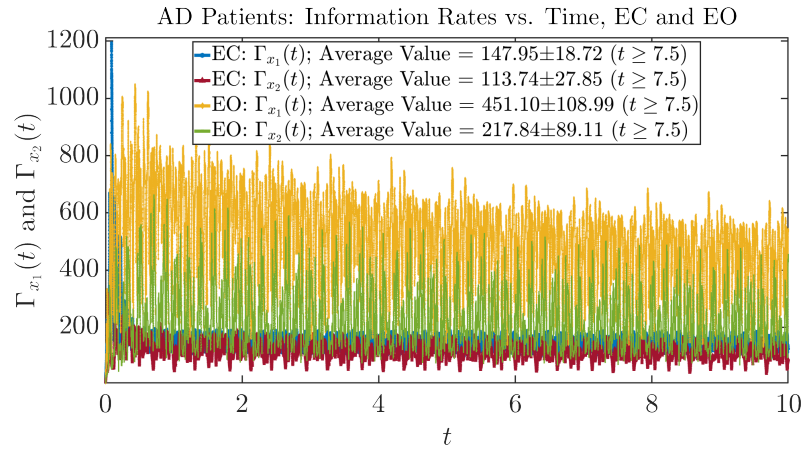
(**b**) AD patients.

Figure 5. Information rates along time of CTL and AD subjects.

Interestingly, when AD patients open their eyes, both $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ are increasing instead of decreasing, as shown in Figure 5b. This might be interpreted as that the incoming visual information received when eyes are opened is in fact "stimulating" the neural information processing activities of the corresponding brain regions, which might be impaired or damaged by the relevant mechanism of Alzheimer's disease (AD).

In Figure 5a,b, we annotate the mean and standard deviation for $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ after $t \geq 7.5$ in the legend, because as mentioned above, the PDFs of this time range reflect longer-term temporal characteristics, and hence, the corresponding $\Gamma_{x_1}(t \geq 7.5)$ and $\Gamma_{x_2}(t \geq 7.5)$ should reflect more reliable and robust features of neural information processing activities of the corresponding brain regions. Therefore, meaningful statistics will require collecting samples of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ in this time range, for which the results are shown in the section below.

3.3.2. Empirical Probability Distribution (for $T \geq 7.5$)

The statistics of $\Gamma_{x_1}(t \geq 7.5)$ and $\Gamma_{x_2}(t \geq 7.5)$ can be further and better visualized using empirical probability distributions of them, as shown in Figure 6. Again, we use histogram-based density estimation with Rice's rule, and since the time interval Δt for estimating PDFs and computing $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ is 10^{-4} (whereas the time-step size dt for simulating the SDE model is 10^{-6}), we collected 24,999 samples of both $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ for $7.5 \leq t < 10$, and hence, the number of bins following Rice's rule is rounded to 58. Figure 6 confirms the observation in the previous section, while it also better visualizes the sample standard deviation in the shapes of the estimated PDFs, indicating that the PDFs of both $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ are narrowed down when healthy subjects open their eyes but are widened when AD patients do so.

As a comparison, we also calculate more traditional/established information-theoretic measure, namely, the Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$, and estimate their empirical probability distributions in the same manner as we do for information rates, as shown in Figure 7.

One can see that the empirical distributions of differential entropy $h(x_1(t))$ and $h(x_2(t))$ are not able to make clear distinction between EC and EO conditions, especially for AD patients. This may be better summarized in Table 4, comparing the mean and standard deviation values of information rate vs. differential entropy for the four cases. Therefore, the information rate is a superior measure for quantifying the non-stationary time-varying dynamical changes in EEG signals when switching between EC and EO states and is a better and more reliable reflection of neural information processing in the brain.

Empirical Prob. Distributions of Information Rates $\Gamma_{x_1}(t \geq 7.5)$ and $\Gamma_{x_2}(t \geq 7.5)$
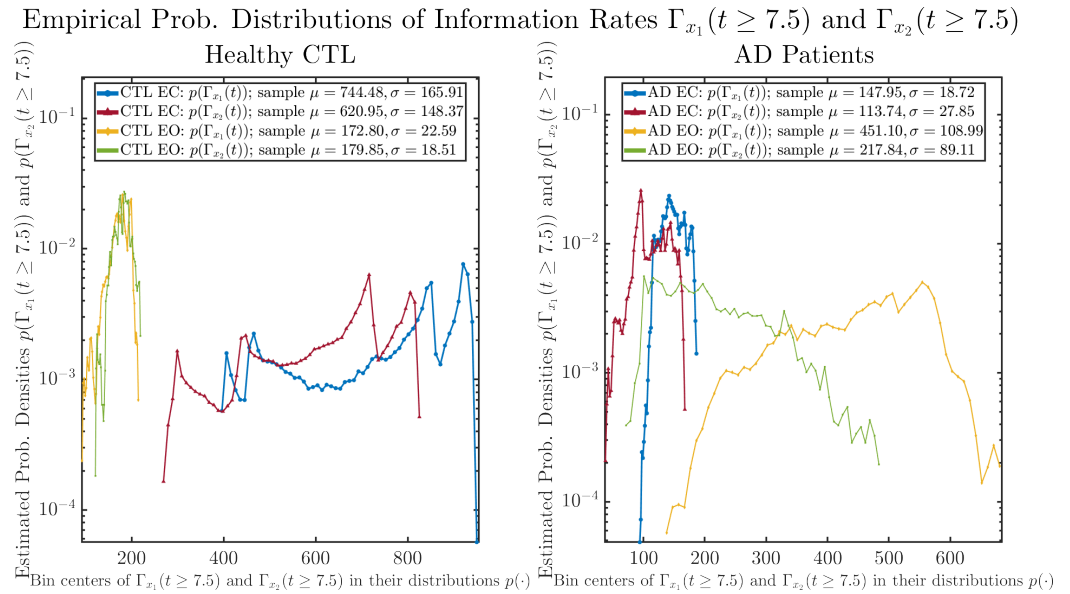
Figure 6. Empirical probability distributions of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ $(t \geq 7.5)$.
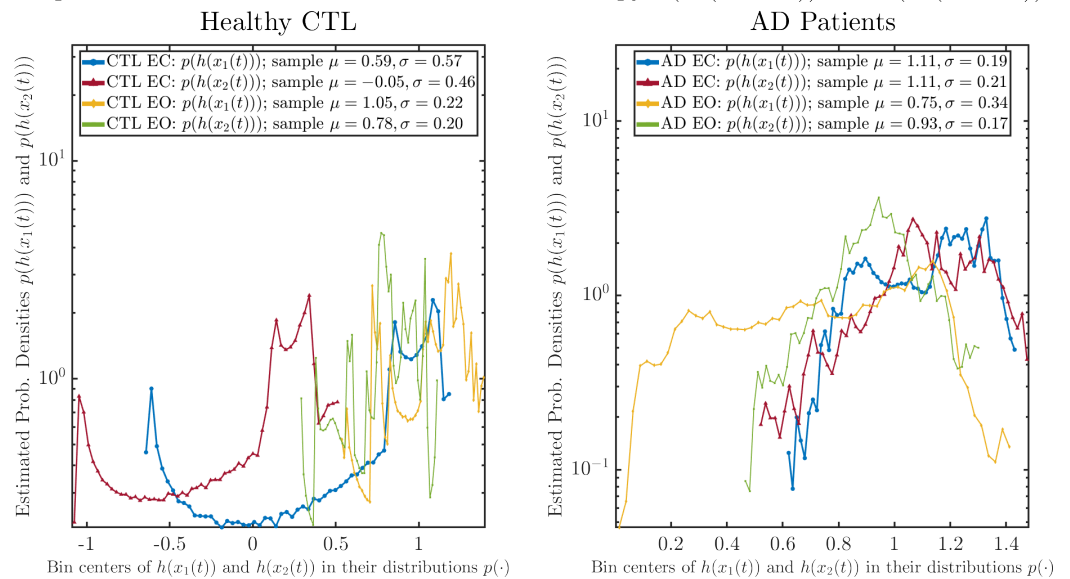
Figure 7. Empirical probability distributions of differential entropy $h(x_1(t))$ and $h(x_2(t))$ $(t \geq 7.5)$.

Table 4. Mean and standard deviation values $(\mu \pm \sigma)$ of information rates $\Gamma_{x_1}(t)$ & $\Gamma_{x_2}(t)$ vs. differential entropy $h(x_1(t))$ & $h(x_2(t))$ $(t \geq 7.5)$.

	CTL EC	CTL EO	AD EC	AD EO
$\Gamma_{x_1}(t)$	744.48 ± 165.91	172.80 ± 22.59	147.95 ± 18.72	451.10 ± 108.99
$\Gamma_{x_2}(t)$	620.95 ± 148.37	179.85 ± 18.51	113.74 ± 27.85	217.84 ± 89.11
$h(x_1(t))$	0.59 ± 0.57	1.05 ± 0.22	1.11 ± 0.19	0.73 ± 0.34
$h(x_2(t))$	-0.05 ± 0.46	0.78 ± 0.20	1.11 ± 0.21	0.93 ± 0.17

3.3.3. Phase Portraits (for $T \geq 7.5$)

In addition to empirical statistics of information rates for $t \geq 7.5$ in terms of estimated probability distributions, one can also visualize the temporal dynamical features of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ combined using phase portraits, as shown in Figure 8. Notice that when healthy

subjects open their eyes, the fluctuation ranges of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ shrink by roughly 5-fold, whereas when AD patients open their eyes, the fluctuation ranges are enlarged.
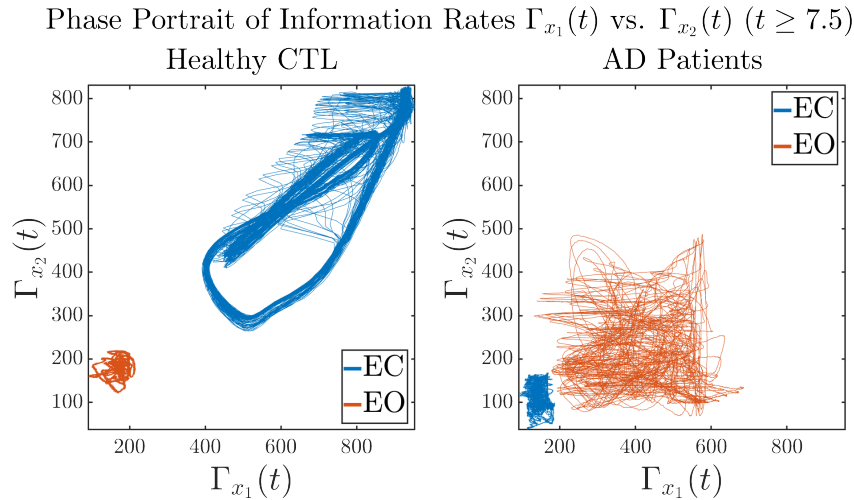
Figure 8. Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$).

Moreover, when plotting EC and EO of healthy subjects separately in Figure 9a to zoom into the ranges of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ for EO, one can also see that the phase portrait of EO exhibits a fractal-like pattern, whereas the phase portrait of EC exhibits more regular dynamical features, including an overall trend of fluctuating between bottom left and top right, indicating that the $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ are somewhat synchronized, which could be explained by the strong coupling coefficients in Table 1 of healthy subjects. Contrarily, for AD patients, the phase portraits of EC and EO both exhibit fractal-like patterns in Figure 9b.

(**a**) Health CTL subjects.   (**b**) AD patients.
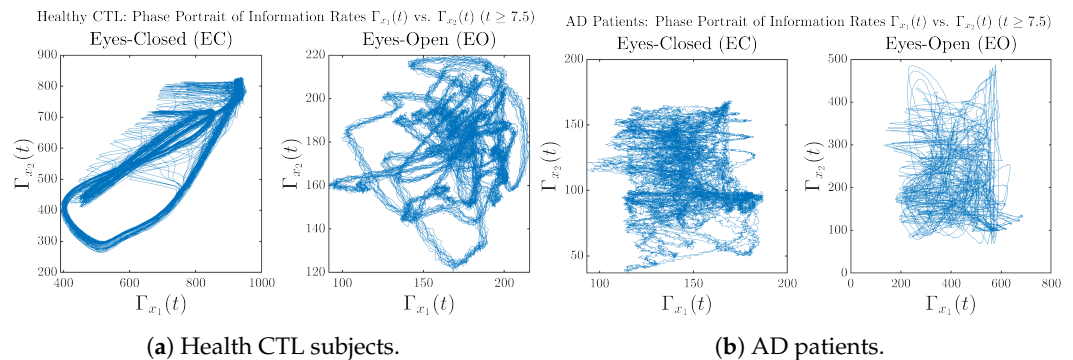
Figure 9. Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Same as at the end of Section 3.3.2, as a comparison, we also visualize the phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ in Figures A52d and A56 in Appendix B.4.3, where one can see that it is hard to distinguish the phase portraits of $h(x_1(t))$ vs. $h(x_2(t))$ of AD EC from those of AD EO, as they are qualitatively the same or very similar. Contrarily, in Figure 8, the fluctuation ranges of phase portraits of $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ are significantly enlarged when AD patients open their eyes. Therefore, this reconfirms our claim at the end of Section 3.3.2 that the information rate is a superior measure than differential entropy in quantifying the dynamical changes in EEG signals when switching between EO and EO states and is a better and more reliable reflection of neural information processing in the brain.

3.3.4. Power Spectra (for $T \geq 7.5$)

Another perspective to visualize the dynamical characteristics of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ is by using power spectra, i.e., the absolute values of (fast) Fourier transforms of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$, as shown in Figure 10. Frequency-based analyses will not make much sense if the

signals or time series of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ have non-stationary time-varying effects, and this is why we only consider time range $t \geq 7.5$ for $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$, when the time evolution patterns of $p(x_1, t)$ and $p(x_2, t)$ almost stop changing as shown in Figure 3 (and especially in Figure 3a,c).
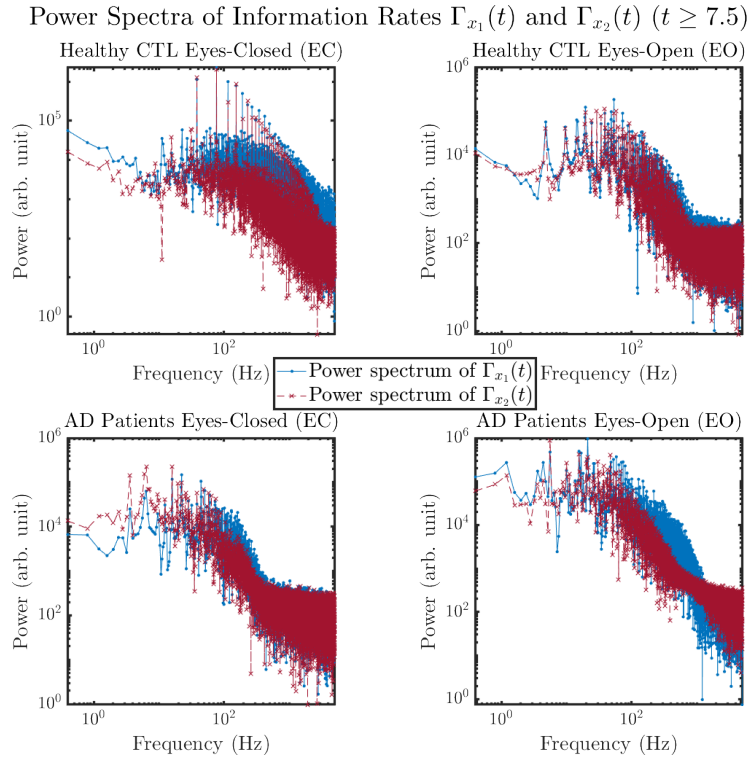
Figure 10. Power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$).

The power spectra of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ also exhibit a clear distinction between EC and EO for CTL and AD subjects. Specifically, the power spectra of $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ can be fit by power law for frequencies between $\sim$100 Hz to $\sim$1000 Hz (the typical sampling frequency of experimental EEG signals is 1000 Hz, whereas most of brain wave's/neural oscillations' frequencies are below 100 Hz). From Figure 11a, one can see that power law fit exponents (quantifying how fast the power density decreases with increasing frequency) of $\Gamma_{x_1}(t)$'s and $\Gamma_{x_2}(t)$'s power spectra are largely reduced when healthy subjects open their eyes, which indicates that the strength of noise in $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ decreases significantly when switching from EC to EO. Contrarily, for AD patients as shown in Figure 11b, the power law fit exponents of $\Gamma_{x_1}(t)$'s and $\Gamma_{x_2}(t)$'s power spectra increase significantly and slightly, respectively, indicating that the strength of noise in $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ increases when switching from EC to EO.

(**a**) Health CTL subjects. (**b**) AD patients.
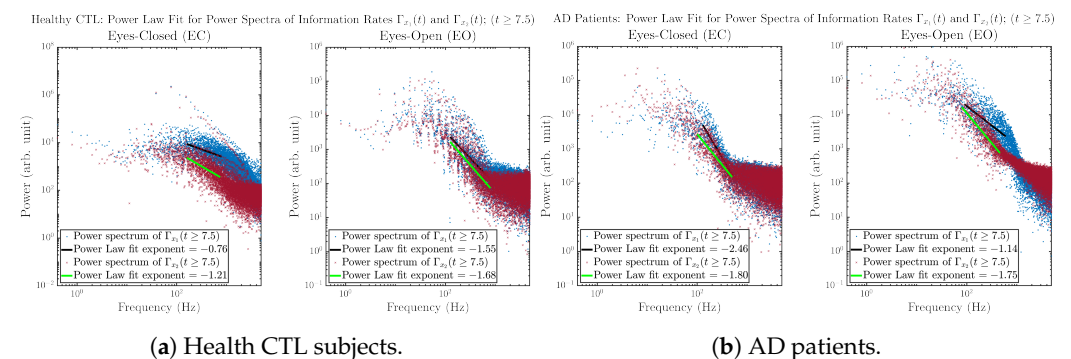
Figure 11. Power law fit for power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

3.4. Causal Information Rates $\Gamma_{x_2 \to X_1}(T), \Gamma_{x_1 \to X_2}(T)$, and Net Causal Information Rates $\Gamma_{x_2 \to X_1}(T) - \Gamma_{x_1 \to X_2}(T)$

The notion of causal information rate was introduced in Ref. [25], which quantifies how one signal instantaneously influences another signal's information rate. A comparable measure of causality is transfer entropy; however, as shown in Appendix B.6, our calculation results of transfer entropy are too spiky/noisy to reliably quantify causality, and hence, the results are only included in Appendix as a comparison, which we will discuss at the end of this section. Nevertheless, similar to net transfer entropy, one can calculate the net causal information rate, e.g., $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$, signifying the net causality measure from signal $x_2(t)$ to $x_1(t)$. Since $\dot{x}_2(t)$ is the only variable that is directly affected by random noise in the stochastic oscillator model Equation (2), we calculate $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ for the net causal information rate of the coupled oscillator's signal $x_2(t)$ influencing $x_1(t)$.

Notice that for stochastic coupled oscillator model Equation (2), causal information rates $\Gamma_{x_2 \to x_1}(t)$ and $\Gamma_{x_1 \to x_2}(t)$ will reflect how strongly the two oscillators are directionally coupled or causally related. Since signals $x_1(t)$ and $x_2(t)$ are the results of neural activities in the corresponding brain regions, the causal information rates can be used to measure connectivities among different regions of the brain.

3.4.1. Time Evolution

Similar to Section 3.3.1, we also visualize the time evolution of causal information rates $\Gamma_{x_2 \to x_1}(t)$ and $\Gamma_{x_1 \to x_2}(t)$ in Figure 12a,b for CTL subjects and AD patients, respectively.

For both CTL and AD subjects, $\Gamma_{x_2 \to x_1}(t)$ and $\Gamma_{x_1 \to x_2}(t)$ both decrease when changing from EC to EO, except for AD subjects' $\Gamma_{x_2 \to x_1}(t)$ increasing on average. On the other hand, the net causal information rate $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ changes differently: when healthy subjects open their eyes, it increases and changes from significantly negative on average to slightly positive on average, whereas for AD patients, it increases from almost zero on average to significantly positive on average without net directional change. A possible interpretation might be that, when healthy subjects open their eyes, the brain region generating the signal $x_2(t)$ becomes more sensitive to the noise, causing it to influence $x_1(t)$ more compared to the eyes-closed state.
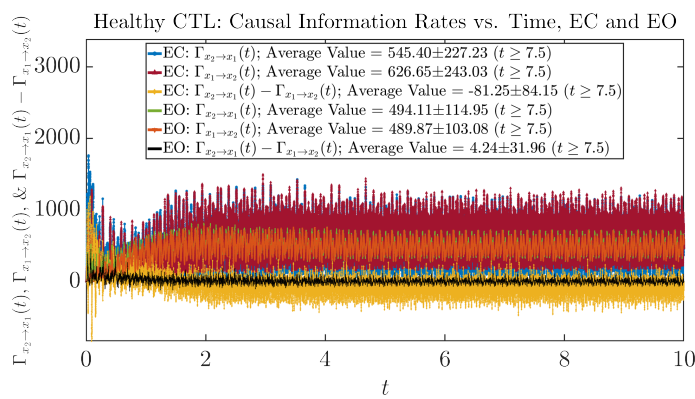
(**a**) health CTL subjects.

Figure 12. *Cont.*
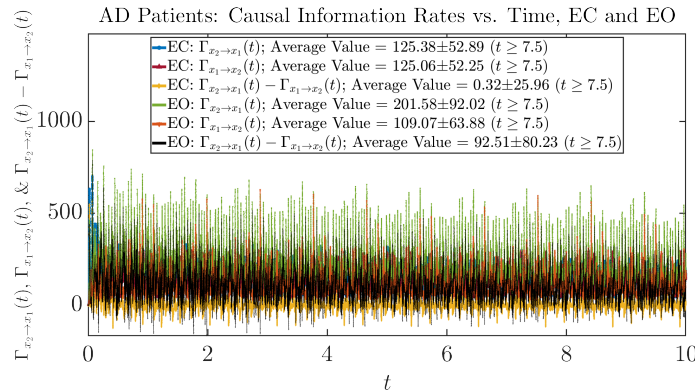
(**b**) AD patients.

Figure 12. Causal information rates along time of CTL and AD subjects.

3.4.2. Empirical Probability Distribution (for $T \geq 7.5$)

Similar to Section 3.3.2, we also estimate the empirical probability distributions of $\Gamma_{x_2 \to x_1}(t), \Gamma_{x_1 \to x_2}(t)$ and $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ to better visualize their statistics in Figure 13a. In particular, we plot the empirical probability distributions of $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ for both healthy and AD subjects with both EC and EO conditions together in Figure 13b, in order to better visualize and compare net causal information rates' changes when CTL and AD subjects open their eyes. It can be seen that the estimated PDF of $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ shrinks its width in shape when healthy subjects open their eyes. Combining with the observation that the magnitude of sample mean of $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ is close to 0 for healthy subjects with eyes opened, a possible interpretation might be that the directional connectivity between brain regions generating signals $x_1(t)$ and $x_2(t)$ is reduced to almost zero, either by incoming visual information received by opened eyes or due to the brain region generating signal $x_2(t)$ becoming more sensitive to noise when eyes are opened. Contrarily, the estimated PDF of $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ for AD patients qualitatively change in an inverse direction to become widened in shape.

As mentioned earlier, as a comparison, we also calculate more traditional/established information-theoretic measure of causality, i.e., transfer entropy (TE), and estimate their empirical probability distributions in the same manner as we do for causal information rates, as shown in Figure 14.

One can see that the empirical distributions of transfer entropy $TE_{x_2 \to x_1}(t)$ and $TE_{x_1 \to x_2}(t)$, as well as net transfer entropy $TE_{x_2 \to x_1}(t) - TE_{x_1 \to x_2}(t)$ are not able to make clear distinction between EC and EO conditions, especially for AD patients in terms of net transfer entropy. This may be better summarized in Table 5, comparing the mean and standard deviation values of causal information rate vs. transfer entropy for the four cases.

Table 5. Mean and standard deviation values ($\mu \pm \sigma$) of causal information rates $\Gamma_{x_2 \to x_1}(t), \Gamma_{x_1 \to x_2}(t)$ and net causal information rates $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$ vs. transfer entropy $TE_{x_2 \to x_1}(t), TE_{x_1 \to x_2}(t)$ and net transfer entropy $TE_{x_2 \to x_1}(t) - TE_{x_1 \to x_2}(t)$ ($t \geq 7.5$).

	CTL EC	CTL EO	AD EC	AD EO
$\Gamma_{x_2 \to x_1}(t)$	545.40 ± 227.23	494.11 ± 114.95	125.38 ± 52.89	201.58 ± 92.02
$\Gamma_{x_1 \to x_2}(t)$	626.65 ± 243.03	489.87 ± 103.08	125.06 ± 52.25	109.07 ± 63.88
$\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$	-81.25 ± 84.15	4.24 ± 31.96	0.32 ± 25.96	92.51 ± 80.23
$TE_{x_2 \to x_1}(t)$	0.011 ± 0.012	0.015 ± 0.0068	0.0067 ± 0.0043	0.0086 ± 0.0062
$TE_{x_1 \to x_2}(t)$	0.013 ± 0.013	0.011 ± 0.0036	0.004 ± 0.0019	0.0046 ± 0.0028
$TE_{x_2 \to x_1}(t) - TE_{x_1 \to x_2}(t)$	-0.0018 ± 0.015	0.0038 ± 0.0062	0.0027 ± 0.0049	0.004 ± 0.0068
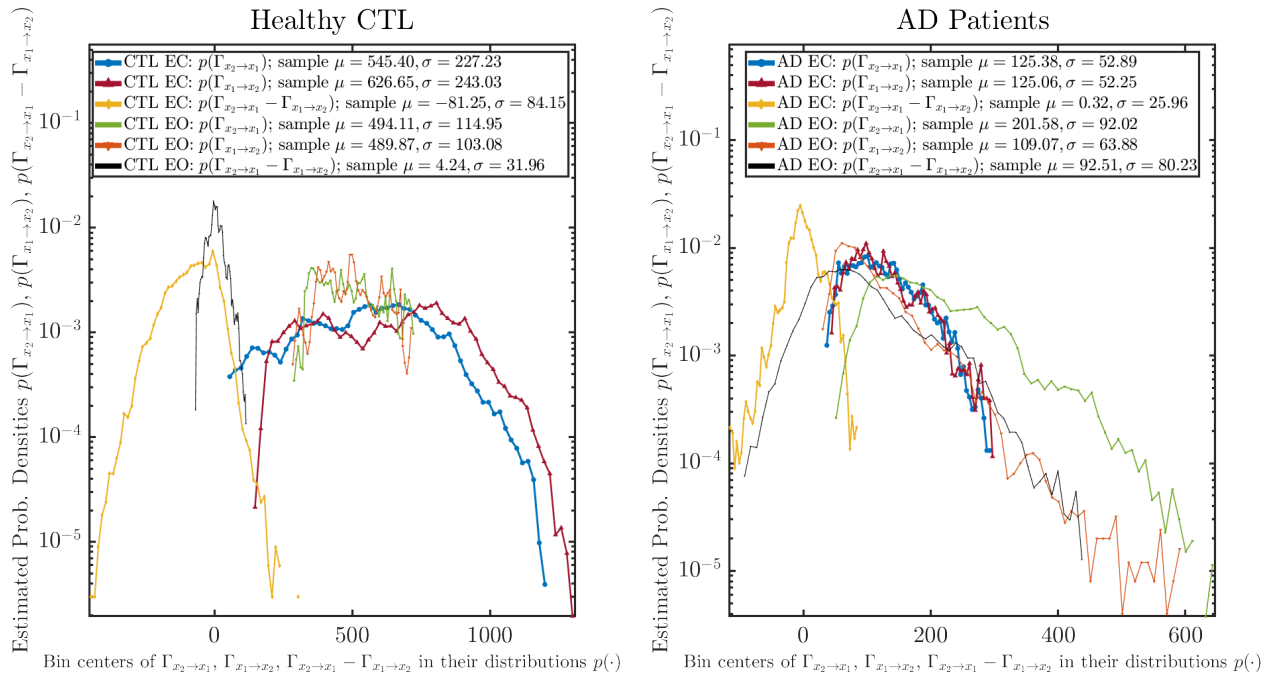
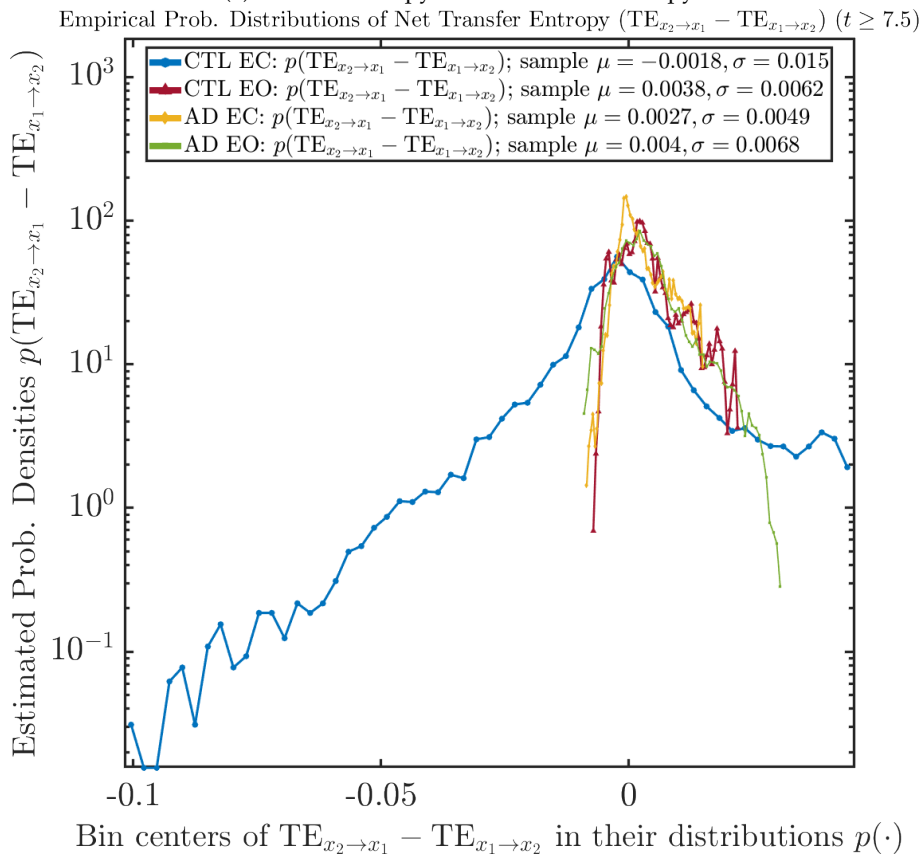(**a**) Causal information rates and net causal information rates.
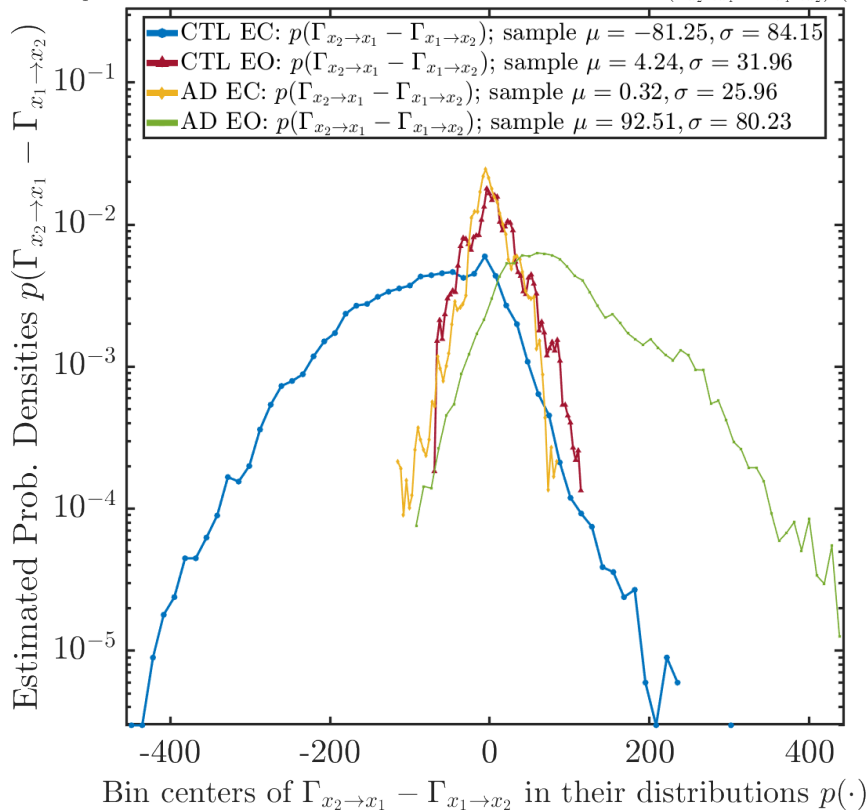
(**b**) Net causal information rates only.

Figure 13. Empirical probability distributions of causal information rates and net causal information rates ($t \geq 7.5$).
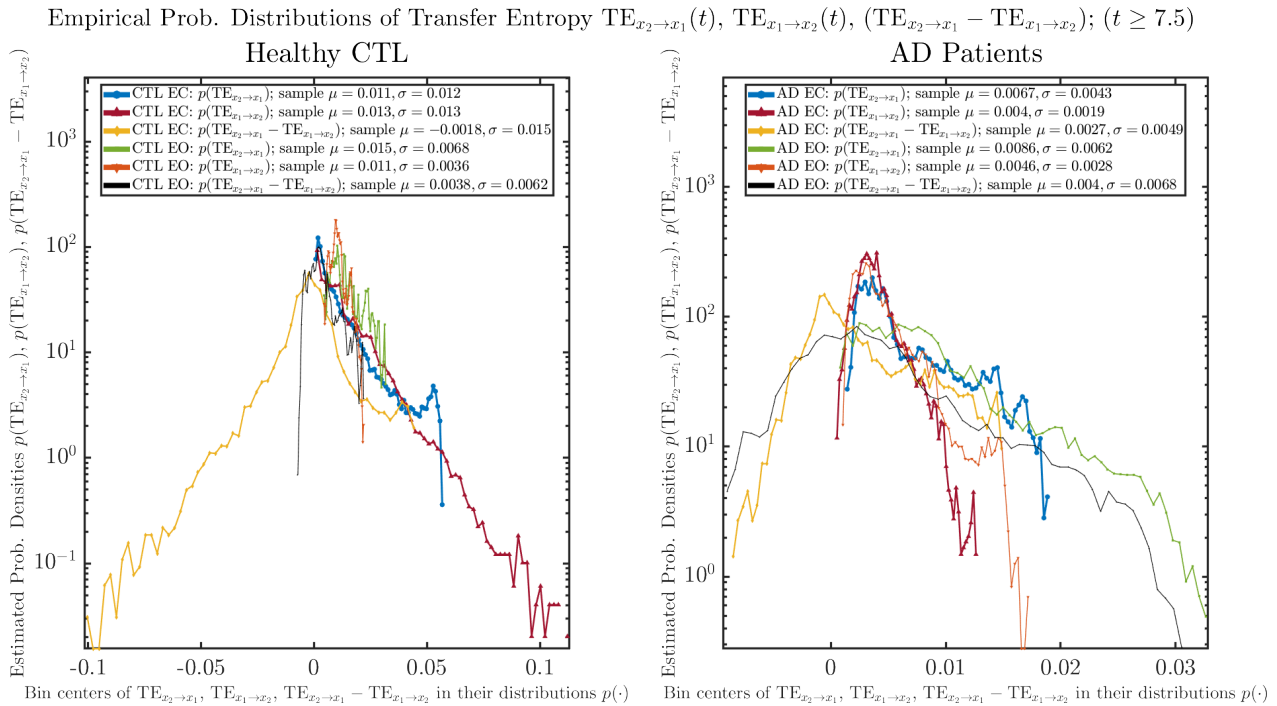
(**a**) Transfer entropy and net transfer entropy.

(**b**) Net transfer entropy only.

Figure 14. Empirical probability distributions of transfer entropy and net transfer entropy ($t \geq 7.5$).

Moreover, the magnitude of numeric values of transfer entropy and net transfer entropy is $\sim 10^{-2}$ or $\sim 10^{-3}$, which is too close to zero, making it too noise-like or unreliable to quantify causality. Therefore, the causal information rate is a much superior measure than transfer entropy in quantifying causality, and since causal information rate quantifies

how one signal instantaneously influences another signal's information rate (which is a reflection of neural information processing in corresponding brain region), it can be used to measure directional or causal connectivities among different brain regions.

4. Discussion

A major challenge for practical usage of information geometry theoretic measures on real-world experimental EEG signals is that they require a significant amount of data samples to estimate the probability density functions. For example, in this work, we simulated 2×10^7 trajectories or sample paths of the stochastic nonlinear coupled oscillator models, such that at any time instance, we always have a sufficient amount of data samples to accurately estimate the time-dependent probability density functions with a histogram-based approach. This is usually not possible for experimental EEG signals which often contain only one trajectory for each channel, and one has to use a sliding window-based approach to collect data samples for histogram-based density estimation. This approach implicitly assumes that the EEG signals are stationary within each sliding time window, and hence, one has to balance between the sliding time window's length and number of available data samples, in order to account for non-stationarity while still having enough data samples to accurately and meaningfully estimate the time-dependent probability densities. And therefore, this approach will not work very well if the EEG signals exhibit severely non-stationary time-varying effects, requiring a very short length of sliding windows, which will contain too few data samples.

An alternative approach to overcome this issue is using kernel density estimation to estimate the probability density functions, which usually requires a much smaller number of data samples while still being able to approximate the true probability distribution with acceptable accuracy. However, this approach typically involves a very high computational cost, limiting its practical use for many cases such as computational resource-limited scenarios. A proposed method to avoid this is using the Koopman operator theoretic framework [37,38] and its numerical techniques applicable to experimental data in a model-free manner, since the Koopman operator is the left-adjoint of the Perron–Frobenious operator evolving the probability density functions in time. This exploration will be left for our future investigation.

5. Conclusions

In this work, we explore information geometry theoretic measures to characterize neural information processing from EEG signals simulated by stochastic nonlinear coupled oscillator models. In particular, we utilize information rates to quantify the time evolution of probability density functions of simulated EEG signals and utilize causal information rates to quantify one signal's instantaneous influence on another signal's information rate. The parameters of the stochastic nonlinear coupled oscillator models of EEG were fine tuned for both healthy subjects and AD patients, with both eyes-closed and eyes-open conditions. By using information rates and causal information rates, we find significant and interesting distinctions between healthy subjects and AD patients when they change their eyes' open/closed status. These distinctions may be further related to differences in neural information processing activities of the corresponding brain regions (for information rates) and to differences in connectivities among these brain regions (for causal information rates).

Compared to more traditional or established information-theoretic measures such as differential entropy and transfer entropy, our results show that information geometry theoretic measures such as information rate and causal information rate are superior to their more traditional counterparts, respectively (information rate vs. differential entropy, and causal information rate vs. transfer entropy). Since information rates and causal information rates can be applied to experimental EEG signals in a model-free manner, and they are capable of quantifying non-stationary time-varying effects, nonlinearity, and non-Gaussian stochasticity presented in real-world EEG signals, we believe that these information geometry theoretic measures can become an important and powerful tool-

set for both understanding neural information processing in the brain and diagnosis of neurological disorders such as Alzheimer's disease in this work.

Abbreviations

The following abbreviations are used in this manuscript:

PDF	probability density function
SDE	stochastic differential equation
IC	Initial Conditions (in terms of initial Gaussian distributions)
CTL	healthy control (subjects)
AD	Alzheimer's disease
EC	eyes-closed
EO	eyes-open
TE	transfer entropy

Appendix A. Finer Details of Numerical Estimation Techniques

Recall from Equation (7) that the squared information rate is

$$\Gamma_x(t)^2 = \int dx \, p(x,t)[\partial_t \ln p(x,t)]^2 = 4 \int dx \left[\partial_t \sqrt{p(x,t)} \right]^2, \tag{A1}$$

where the partial time derivative and integral can be numerically approximated using discretization, i.e., $\partial_t \sqrt{p(x,t)} \approx \frac{1}{(\Delta t)}(\sqrt{p(x,t+\Delta t)} - \sqrt{p(x,t)})$ and $\int dx \, f(x) \approx \sum_i \Delta x_i f(x_i)$, respectively, where for brevity and if no ambiguity arises, the summation over index i is often omitted and replaced by x itself as $\sum_x \Delta x \, f(x)$, and the symbol x serves as both the index of summation (e.g., the x-th interval with length Δx) and the actual value x in $f(x)$.

A common technique to improve the approximation of integral by finite summation is the trapezoidal rule $\int dx \, f(x) \approx \sum_i \Delta x_i \frac{f(x_{i-1}) + f(x_i)}{2}$, which will be abbreviated as $\sum_x^{\text{Trapz.}} \Delta x \, f(x)$ to indicate that the summation is following the trapezoidal rule imposing a 1/2 weight/factor on the first and last summation terms (corresponding to the lower and upper bounds of the integral). Similarly, we use $\sum_{x,y}^{\text{Trapz.2D}} \Delta x \Delta y \, f(x,y)$ to denote

a 2D trapezoidal approximation of the double integral $\int dx\, dy\, f(x,y)$, where different weights (1/4 or 1/2) will be applied to the "corner"/boundary terms of the summation. Meanwhile, to distinguish regular summation from trapezoidal approximation, we use the notation $\sum_x^{\text{naive}} \Delta x\, f(x)$ to signify a regular summation as a more naive approximation of the integral.

The PDF $p(x,t)$ is numerically estimated using a histogram with Rice's rule applied, i.e., the number of bins is $n_{\text{bins}} = \left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor$ (with uniform bin width $\Delta x = \frac{\text{range of samples' values}}{n_{\text{bins}}}$), which is rounded towards zero to avoid overestimating the number of bins needed. And for joint PDF $p(x_i, t_i; x_j, t_j)$, since the bins are distributed in a 2D plane of (x_i, x_j), the number of bins in each dimension is rounded to $\left\lfloor \sqrt{\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor} \right\rfloor$ (and similarly for 3D joint probability as in the transfer entropy calculation, the number of bins in each dimension is rounded to $\left\lfloor \sqrt[3]{\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor} \right\rfloor$). Combining all of the above, the information rate's square will be approximated by

$$\Gamma_x(t)^2 = 4 \int dx \left[\partial_t \sqrt{p(x,t)} \right]^2 \stackrel{\text{Trapz.}}{\approx} 4 \sum_x \frac{\Delta x}{[\Delta t]^2} \left[\sqrt{p(x,t+\Delta t)} - \sqrt{p(x,t)} \right]^2, \quad \text{(A2)}$$

where the bin width Δx can be moved into the square root and multiplied with the PDF to get the probability (mass) of finding a data sample in the x-th bin, which is estimated as $\frac{n_{\text{sample inside } x\text{-th bin}}}{n_{\text{sample}}}$, i.e., the number of data samples inside that bin divided by the number of all data samples (using the relevant functions in MATLAB or Python). The trapezoidal rule imposes a 1/2 factor on the first and last terms of summation, corresponding to the first and last bins.

For the causal information rate $\Gamma_{x_1 \to x_2}(t) \stackrel{\text{def}}{=} \Gamma_{x_2}^*(t) - \Gamma_{x_2}(t)$, the $\Gamma_{x_2}^*(t)$ can be estimated by

$$\Gamma_{x_2}^*(t)^2 = \lim_{t_* \to t^+} \int dx_1\, dx_2\, p(x_2, t_*; x_1, t) [\partial_{t_*} \ln p(x_2, t_*; x_1, t)]^2$$

$$= 4 \lim_{t_* \to t^+} \int dx_1\, dx_2 \left[\partial_{t_*} \sqrt{p(x_2, t_*; x_1, t)} \right]^2 \quad \text{(A3)}$$

$$\approx 4 \sum_{x_2, x_1} \frac{\Delta x_2 \Delta x_1}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t; x_1, t)} - \sqrt{p(x_2, t; x_1, t)} \right]^2,$$

where the number of bins in each of the x_1 and x_2 dimensions is rounded to $\left\lfloor \sqrt{\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor} \right\rfloor$, and the $\Gamma_{x_2}(t)^2$ can be estimated as above as $4 \int dx_2 \left[\partial_t \sqrt{p(x_2,t)} \right]^2 \approx 4 \sum_{x_2} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2,t)} \right]^2$ using regular or trapezoidal summation. However, here for $\Gamma_{x_2}(t)^2$, the number of bins for x_2 must *not* be chosen as $\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor$ following the 1D Rice's rule, which is very critical to avoid insensible or inconsistent estimation of $\Gamma_{x_1 \to x_2}(t)$, for which the reason is explained below.

Consider the quantity $\Gamma_{x_2}^*(t)^2 - \Gamma_{x_2}(t)^2$; theoretically and by definition, the dx_2 can be pulled outside the integral over dx_1 to combine the two integrals into one integral as follows:

$$\Gamma_{x_2}^*(t)^2 - \Gamma_{x_2}(t)^2 = 4 \lim_{t_* \to t^+} \int dx_2 \left\{ \int \left[\partial_{t_*} \sqrt{p(x_2, t_*; x_1, t)} \right]^2 dx_1 - \left[\partial_t \sqrt{p(x_2, t)} \right]^2 \right\}, \quad \text{(A4)}$$

and the corresponding numerical approximations of integrals should be combined as

$$\approx 4 \sum_{x_2} \Delta x_2 \left\{ \sum_{x_1} \left[\sqrt{p(x_2, t+\Delta t; x_1, t)} - \sqrt{p(x_2, t; x_1, t)} \right]^2 \frac{\Delta x_1}{[\Delta t]^2} \right.$$
$$\left. - \frac{\left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2}{[\Delta t]^2} \right\}, \tag{A5}$$

where the sum over x_2 is performed on the same bins for both of the two terms inside the large braces $\{\cdot\}$ above. On the other hand, if one numerically approximates $\Gamma_{x_2}^*(t)^2$ and $\Gamma_{x_2}(t)^2$ separately as

$$\Gamma_{x_2}^*(t)^2 - \Gamma_{x_2}(t)^2 \approx 4 \sum_{x_2, x_1} \frac{\Delta x_2 \Delta x_1}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t; x_1, t)} - \sqrt{p(x_2, t; x_1, t)} \right]^2$$
$$- 4 \sum_{x_2} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2, \tag{A6}$$

then the sum over x_2 in the second term $4\sum_{x_2} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2$ should *still* be performed on the same bins of x_2 for the first term involving the joint PDFs estimated by 2D histograms (i.e., using the square root number of bins $\left\lfloor \sqrt{\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor} \right\rfloor$ of Rice's rule, instead of following the 1D Rice's rule without the square root), even though this second summation term is written as a separate and "independent" term from the first double-summation term. The definition $\Gamma_{x_1 \to x_2}(t) \overset{\text{def}}{=} \Gamma_{x_2}^*(t) - \Gamma_{x_2}(t)$ might result in a misimpression that one can estimate $\Gamma_{x_2}(t)^2 \approx 4\sum_{x_2} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2$ separately by using a Rice's rule's binning method containing $\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor$ bins, while estimating $\Gamma_{x_2}^*(t)^2 \approx 4\sum_{x_2, x_1} \frac{\Delta x_2 \Delta x_1}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t; x_1, t)} - \sqrt{p(x_2, t; x_1, t)} \right]^2$ using the square root of Rice's rule's number of bins $\left\lfloor \sqrt{\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor} \right\rfloor$. Using different bins for x_2 will make it invalid to combine the two summations into one summation over the same x_2's (and hence invalid to combine the two integrals into one integral by pulling out the same dx_2).

Using $\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor$ bins for x_2 will overestimate the value of $\Gamma_{x_2}(t)^2 \approx 4\sum_{x_2} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2$, for example, if there are 1 million samples/data points to estimate the PDFs, then $\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor = 200$ for 1D distribution and $\left\lfloor \sqrt{\left\lfloor 2\sqrt[3]{n_{\text{sample}}} \right\rfloor} \right\rfloor \approx 14$ for 2D joint distribution. Calculating $\Gamma_{x_2}(t)^2 \approx 4\sum_{x_2} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2$ using 200 bins will result in a much larger value than calculating it using 14 bins, which will result in negative values in calculating the causal information rate $\Gamma_{x_1 \to x_2}(t) = \Gamma_{x_2}^*(t) - \Gamma_{x_2}(t)$. When using the same 14 bins of x_2 (for estimating the 2D joint PDF of (x_1, x_2)) to estimate the 1D PDF in $\Gamma_{x_2}(t)^2 \approx 4\sum_{x_2} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2$, all the unreasonable negative values disappear, except for only some isolated negative values remained, which is related to estimating $\Gamma_{x_2}^*(t)^2$ and $\Gamma_{x_2}(t)^2$ using 1D and 2D trapezoidal rules for summations approximating the integrals: if one uses 1D trapezoidal summation for $\Gamma_{x_2}(t)^2 \approx 4\sum_{x_2}^{\text{Trapz.}} \frac{\Delta x_2}{[\Delta t]^2} \left[\sqrt{p(x_2, t+\Delta t)} - \sqrt{p(x_2, t)} \right]^2$, while on the other hand, one blindly and inconsistently uses 2D trapezoidal summation for $\Gamma_{x_2}^*(t)^2 \approx 4\sum_{x_2, x_1}^{\text{Trapz.2D}} \frac{\Delta x_2 \Delta x_1}{[\Delta t]^2} [\sqrt{p(x_2, t+\Delta t; x_1, t)} - \sqrt{p(x_2, t; x_1, t)}]^2$, this will also result in some negative values in computing $\Gamma_{x_1 \to x_2}(t)$

$= \Gamma^*_{x_2}(t) - \Gamma_{x_2}(t)$, because the 2D trapezoidal sum will under-estimate the $\Gamma^*_{x_2}(t)^2$ as compared to the 1D trapezoidal-sum-estimated $\Gamma_{x_2}(t)^2$.

To resolve this inconsistent mixing of 1D and 2D trapezoidal rules, there are two possible methods:

1. Using 2D trapezoidal rule for both $\Gamma^*_{x_2}(t)^2$ and $\Gamma_{x_2}(t)^2$, that is, $\Gamma^*_{x_2}(t)^2 \approx 4\sum^{\text{Trapz.2D}}_{x_2,x_1} \frac{\Delta x_2 \Delta x_1}{[\Delta t]^2}[\sqrt{p(x_2,t+\Delta t;x_1,t)} - \sqrt{p(x_2,t;x_1,t)}]^2$, and $\Gamma_{x_2}(t)^2 \approx 4\sum^{\text{Trapz.}}_{x_2} \frac{\Delta x_2}{[\Delta t]^2}\left[\sqrt{p(x_2,t+\Delta t)} - \sqrt{p(x_2,t)}\right]^2 \approx 4\sum^{\text{Trapz.}}_{x_2} \frac{\Delta x_2}{[\Delta t]^2}[\sqrt{\sum^{\text{Trapz.}}_{x_1} \Delta x_1 p(x_2,t+\Delta t;x_1,t)} - \sqrt{\sum^{\text{Trapz.}}_{x_1} \Delta x_1 p(x_2,t;x_1,t)}]^2$. In other words, when calculating $\Gamma_{x_2}(t)^2$, instead of estimating marginal PDF $p(x_2,t+\Delta t)$ and $p(x_2,t)$ directly by 1D histograms (using the relevant functions in MATLAB or Python), one first estimates the joint PDF $p(x_2,t+\Delta t;x_1,t)$ and $p(x_2,t;x_1,t)$ by 2D histograms and integrates over x_1 by trapezoidal summation on it. This will reduce the value of estimated $\Gamma_{x_2}(t)$, and integrals over both x_1 and x_2 are both estimated by trapezoidal summation.

2. Using the 1D trapezoidal rule for both $\Gamma_{x_2}(t)$ and $\Gamma^*_{x_2}(t)$, that is, $\Gamma_{x_2}(t)^2 \approx 4\sum^{\text{Trapz.}}_{x_2} \frac{\Delta x_2}{[\Delta t]^2}\left[\sqrt{p(x_2,t+\Delta t)} - \sqrt{p(x_2,t)}\right]^2 = 4\sum^{\text{Trapz.}}_{x_2} \frac{\Delta x_2}{[\Delta t]^2}[\sqrt{\sum^{\text{naive}}_{x_1} \Delta x_1 \cdot p(x_2,t+\Delta t;x_1,t)} - \sqrt{\sum^{\text{naive}}_{x_1} \Delta x_1 \cdot p(x_2,t;x_1,t)}]^2$, and $\Gamma^*_{x_2}(t)^2 \approx \frac{4}{[\Delta t]^2}\sum^{\text{Trapz.}}_{x_2} \Delta x_2 \left\{\sum^{\text{naive}}_{x_1} \Delta x_1 \left[\sqrt{p(x_2,t+\Delta t;x_1,t)} - \sqrt{p(x_2,t;x_1,t)}\right]^2\right\}$. In this approach, the marginal PDF $p(x_2,t) = \int p(x_2,t;x_1,t)dx_1$, where the equal sign holds exactly for the regular or naive summation $p(x_2,t) = \sum^{\text{naive}}_{x_1} \Delta x_1 \cdot p(x_2,t;x_1,t)$. This is because the histogram estimation in MATLAB and Python is performed by counting the occurrence of data samples inside each bin, and the probability (mass) is estimated as $\frac{n_{\text{sample}} \text{ inside } x\text{-th bin}}{n_{\text{sample}}}$, and the density is estimated as $\frac{n_{\text{sample}} \text{ inside } x\text{-th bin}}{n_{\text{sample}} \cdot \Delta x}$, where Δx is the width of the x-th bin (and for 2D histogram, this is replaced by bin area $A_{x_1,x_2} = \Delta x_1 \cdot \Delta x_2$), and therefore, summing over x_1 is aggregating the 2D bins of (x_1,x_2) and combining or mixing samples with x_2-values/coordinates in the same x_2-bin (but with x_1-values/coordinates in different x_1-bins) together. In other words, it is always true that $n_{x_2} = \sum^{\text{naive}}_{x_1} n_{x_1,x_2}$, where n_{x_2} is the number of samples inside the x_2-th bin and n_{x_1,x_2} is number of samples inside the (x_1,x_2)-th bin in 2D, and hence, for estimated probability (mass), $\frac{n_{x_2}}{n_{\text{sample}}} = \sum^{\text{naive}}_{x_1} \frac{n_{x_1,x_2}}{n_{\text{sample}}}$, and for estimated PDFs, $\frac{n_{x_2}}{n_{\text{sample}} \cdot \Delta x_2} = \sum^{\text{naive}}_{x_1} \frac{n_{x_1,x_2}}{n_{\text{sample}} \cdot \Delta x_2 \cdot \Delta x_1} \cdot \Delta x_1$, which is why $p(x_2,t) = \sum^{\text{naive}}_{x_1} \Delta x_1 \cdot p(x_2,t;x_1,t)$ holds exactly for numerically estimated marginal and joint PDFs using histograms, which is consistent with the theoretical relation between marginal and joint PDFs $p(x_2,t) = \int p(x_2,t;x_1,t)dx_1$, and this has been numerically verified using the relevant 1D and 2D histogram functions in MATLAB and Python, i.e., by (naively) summing the estimated joint PDF over x_1, and the (naively) summed marginal is exactly the same as the one estimated directly by 1D histogram function. So in this approach, integral over x_1 is estimated by naive summation on x_1, but integral over x_2 is estimated by trapezoidal summation on x_2.

The 1st approach will violate the relation between joint and marginal $p(x_2, t)$ $= \int p(x_2, t; x_1, t) dx_1$, because as explained in the 2nd approach above, when using MATLAB's and Python's 1D and 2D histogram functions, one will always get exactly $p(x_2, t)$ $= \sum_{x_1}^{\text{naive}} \Delta x_1 \cdot p(x_2, t; x_1, t)$ and $p(x_2, t + \Delta t) = \sum_{x_1}^{\text{naive}} \Delta x_1 \cdot p(x_2 + \Delta t, t; x_1, t)$ for naive summation, but *not* for trapezoidal summation over x_1 due to the weights/factors ($\neq 1$) imposed on the "corner"/boundary/first/last summation terms, which is used in the 1st approach. However, the 2nd approach puts different importance or weights on the summation over x_1 as compared to x_2, which might also be problematic, because the original definition is a double integral over x_1 and x_2 without different weights/factors imposed by different summation methods.

To resolve this, we use the regular or naive summations on both x_1 and x_2, which avoids the issues in both the 1st and 2nd approaches, and we find that the numerical difference between the 1st and 2nd approaches and our adopted simply naive summations are really negligible, and because in this work, we are performing empirical statistics on the estimated causal information rates and illustrating the qualitative features of the empirical probability distributions of them, we use our simple naive summations over both x_1 and x_2 when estimating $\Gamma_{x_2}^*(t)^2$ and $\Gamma_{x_2}(t)^2$ in causal information rate $\Gamma_{x_1 \to x_2}(t) = \Gamma_{x_2}^*(t) - \Gamma_{x_2}(t)$.

Appendix B. Complete Results: All Six Groups of Initial Conditions

For completeness, we list the full results of all figures for all six different initial Gaussian distributions listed in Table 3.

Appendix B.1. Sample Trajectories of $x_1(t)$ and $x_2(t)$

Appendix B.1.1. Initial Conditions No.1 (IC1)

(a) IC1: CTL EC $x_1(t)$ **(b)** IC1: CTL EO $x_1(t)$

(c) IC1: CTL EC $x_2(t)$ **(d)** IC1: CTL EO $x_2(t)$
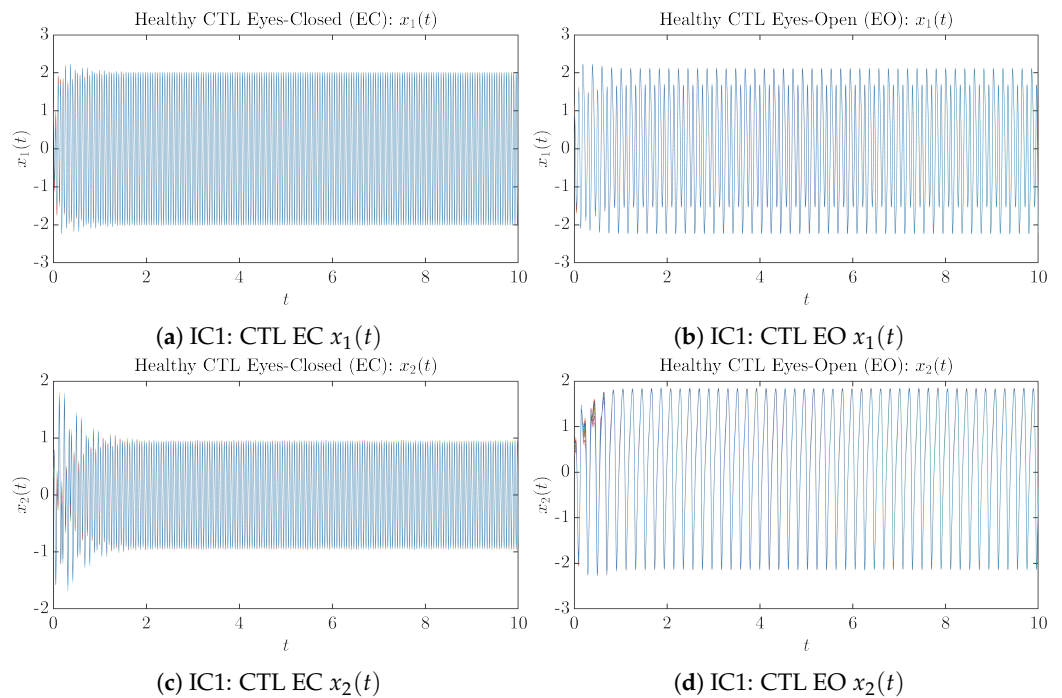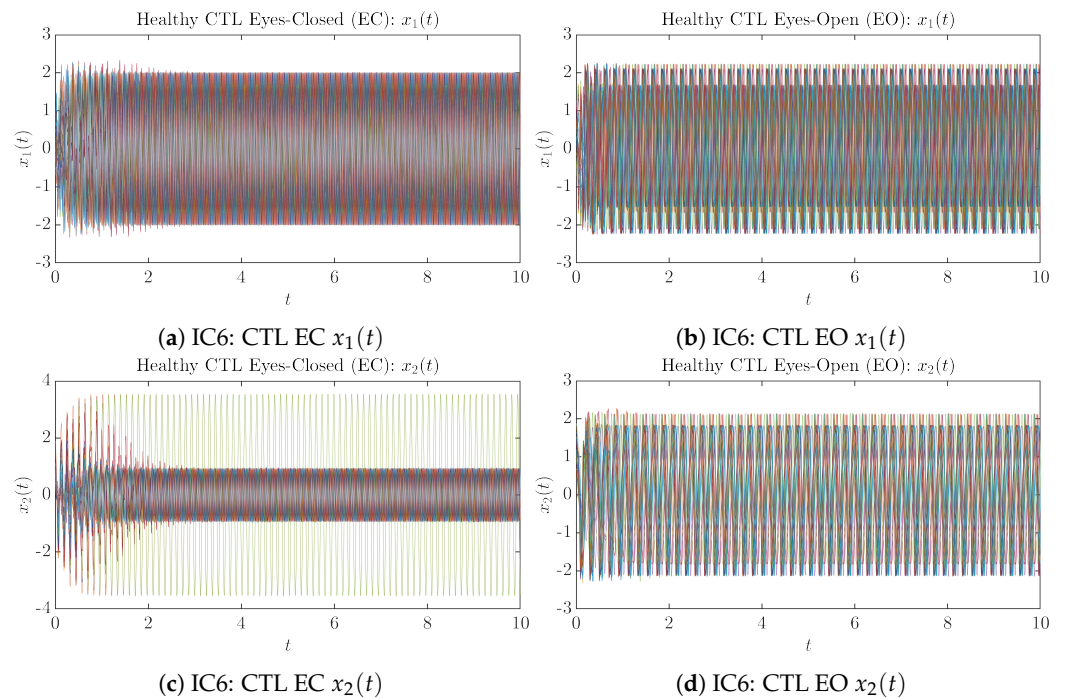
Figure A1. Initial Conditions No.1 (IC1): 50 sample trajectories of healthy CTL subjects. Each single trajectory is labeled by a different color.
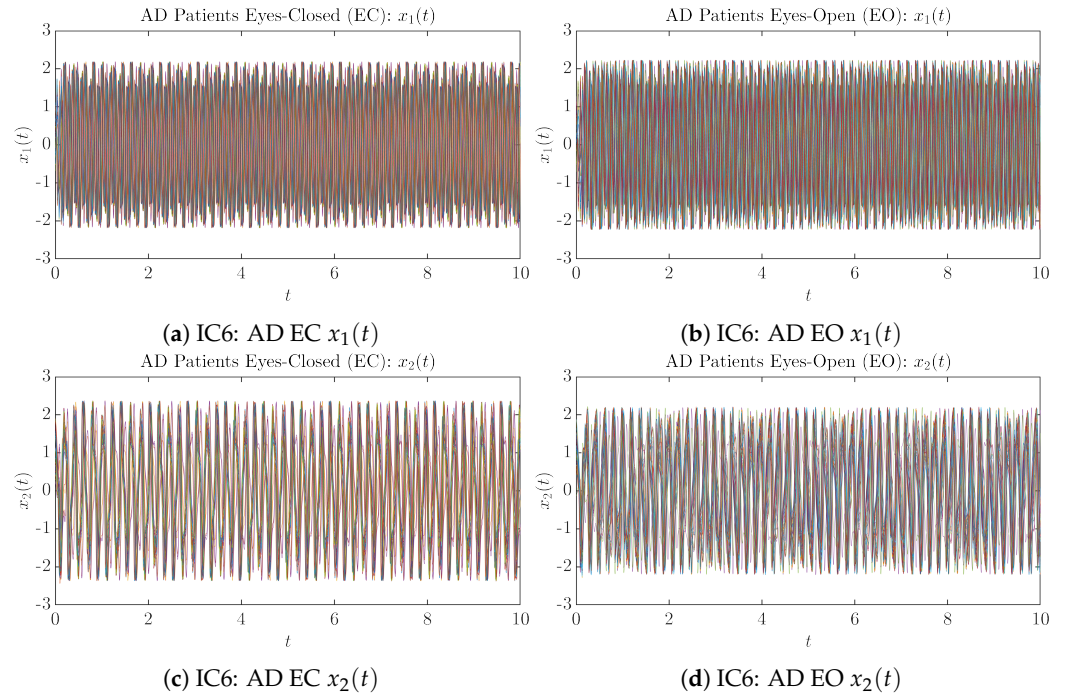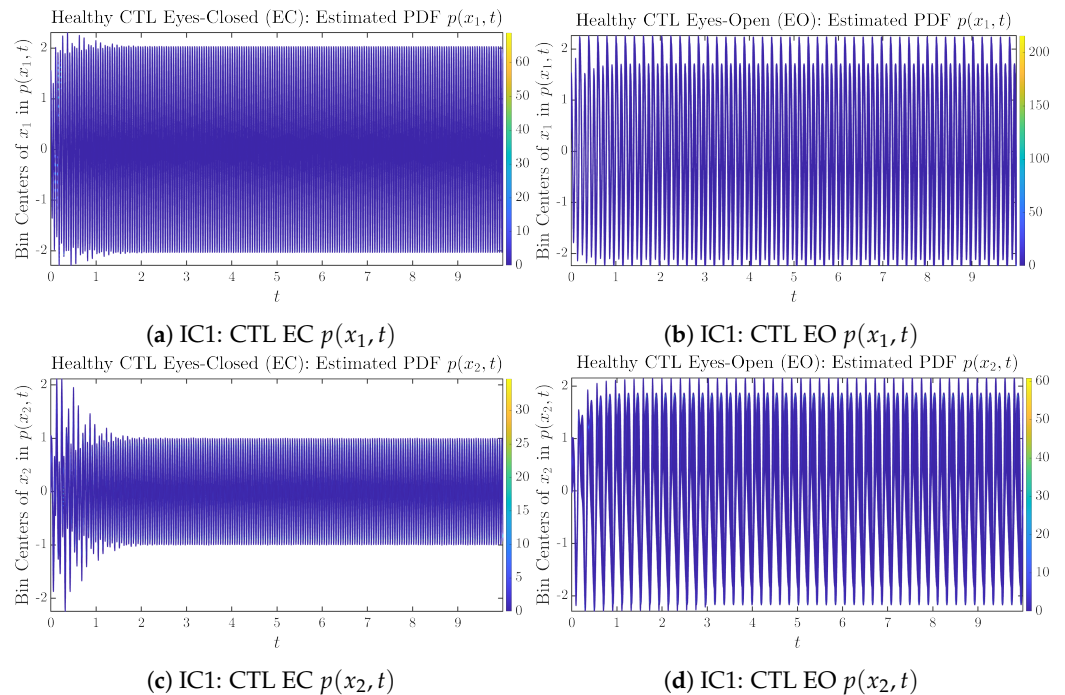
(a) IC1: AD EC $x_1(t)$

(b) IC1: AD EO $x_1(t)$

(c) IC1: AD EC $x_2(t)$

(d) IC1: AD EO $x_2(t)$
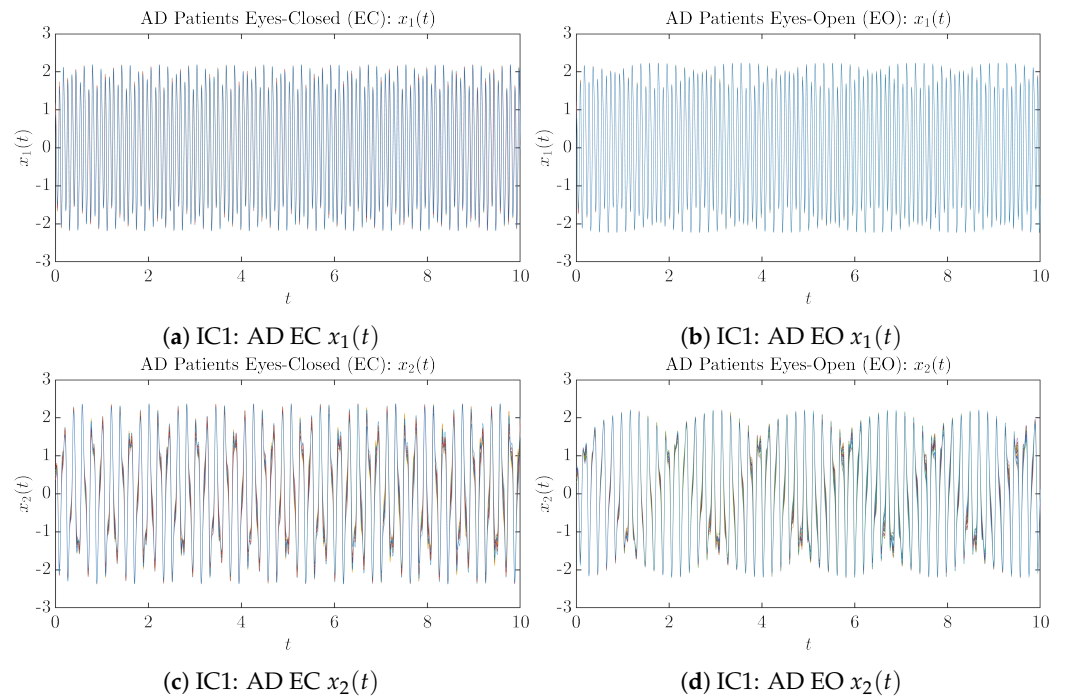
Figure A2. Initial Conditions No.1 (IC1): 50 sample trajectories of AD patients. Each single trajectory is labeled by a different color.

Appendix B.1.2. Initial Conditions No.2 (IC2)

(a) IC2: CTL EC $x_1(t)$

(b) IC2: CTL EO $x_1(t)$

(c) IC2: CTL EC $x_2(t)$

(d) IC2: CTL EO $x_2(t)$
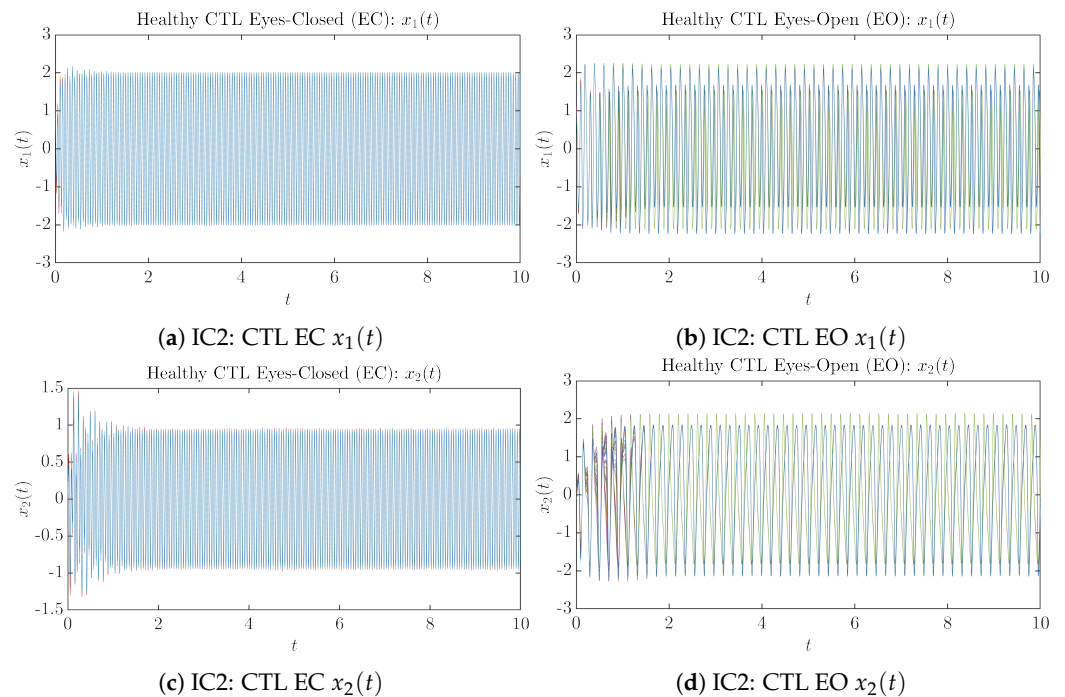
Figure A3. Initial Conditions No.2 (IC2): 50 sample trajectories of healthy CTL subjects. Each single trajectory is labeled by a different color.

(**a**) IC2: AD EC $x_1(t)$

(**b**) IC2: AD EO $x_1(t)$

(**c**) IC2: AD EC $x_2(t)$

(**d**) IC2: AD EO $x_2(t)$
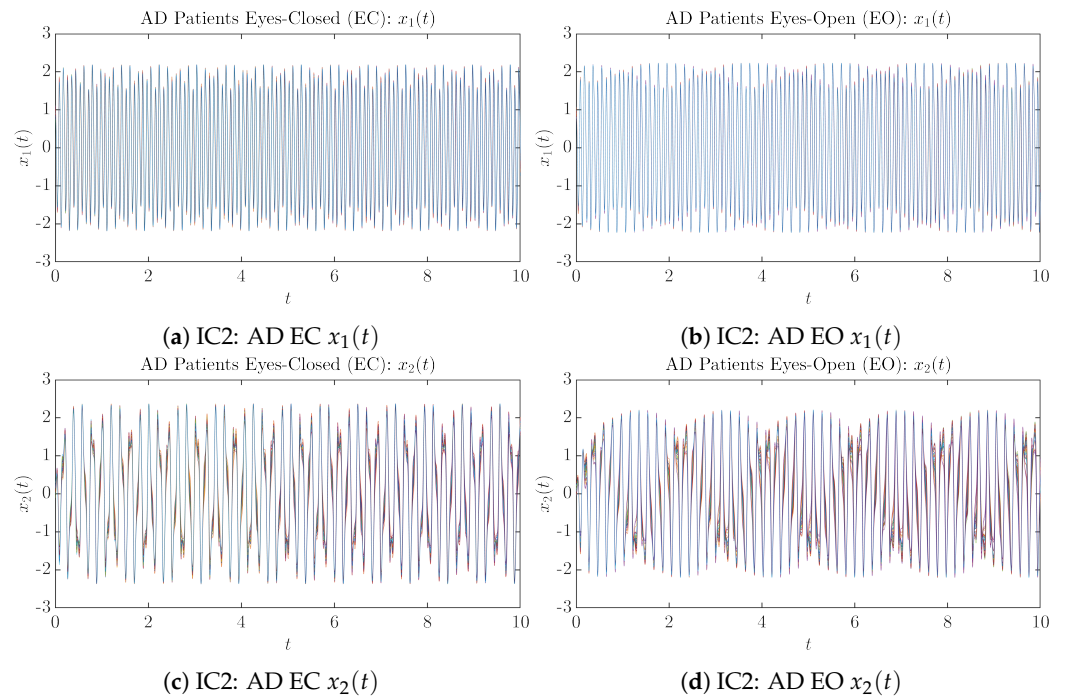
Figure A4. Initial Conditions No.2 (IC2): 50 sample trajectories of AD patients. Each single trajectory is labeled by a different color.

Appendix B.1.3. Initial Conditions No.3 (IC3)

(**a**) IC3: CTL EC $x_1(t)$

(**b**) IC3: CTL EO $x_1(t)$

(**c**) IC3: CTL EC $x_2(t)$

(**d**) IC3: CTL EO $x_2(t)$
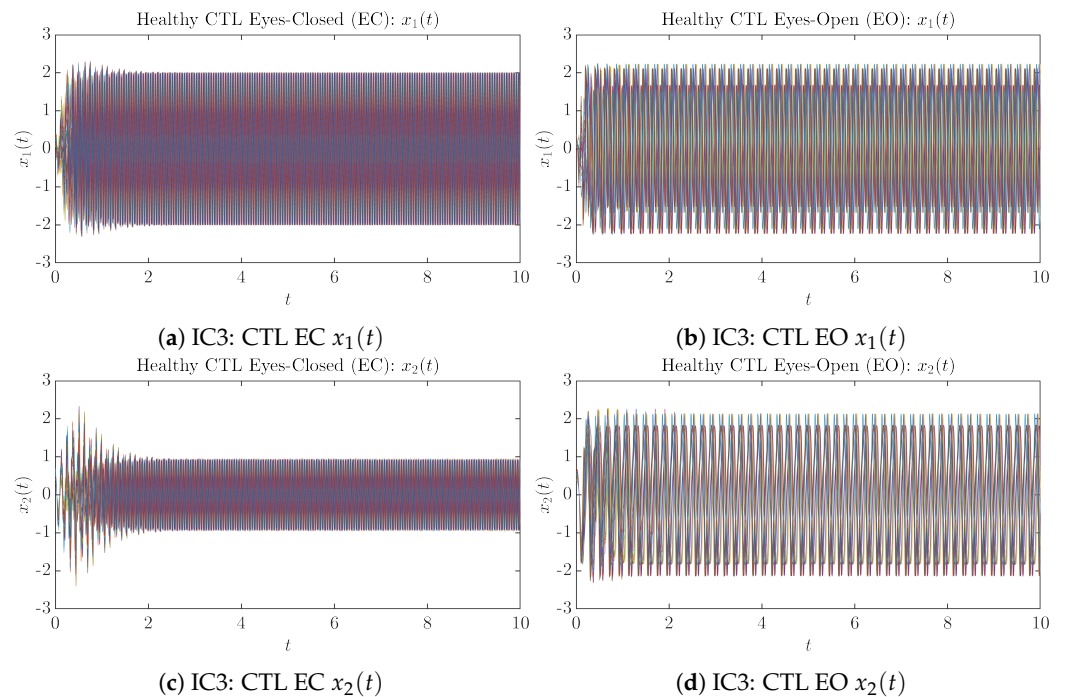
Figure A5. Initial Conditions No.3 (IC3): 50 sample trajectories of healthy CTL subjects. Each single trajectory is labeled by a different color.

(**a**) IC3: AD EC $x_1(t)$

(**b**) IC3: AD EO $x_1(t)$

(**c**) IC3: AD EC $x_2(t)$

(**d**) IC3: AD EO $x_2(t)$
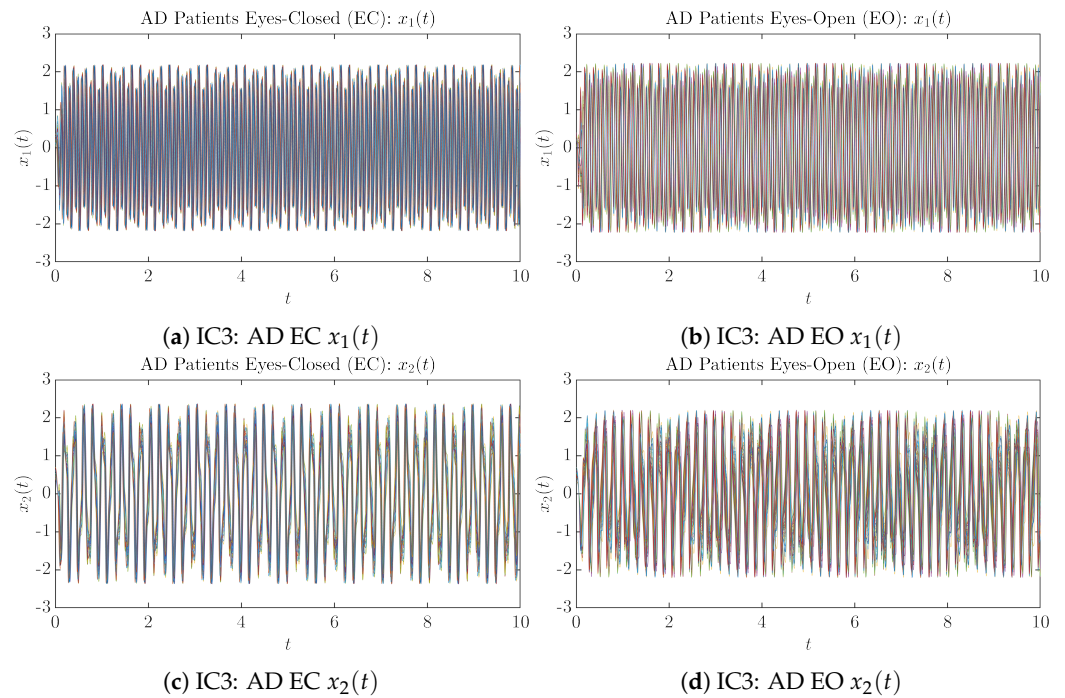
Figure A6. Initial Conditions No.3 (IC3): 50 sample trajectories of AD patients. Each single trajectory is labeled by a different color.

Appendix B.1.4. Initial Conditions No.4 (IC4)

(**a**) IC4: CTL EC $x_1(t)$

(**b**) IC4: CTL EO $x_1(t)$

(**c**) IC4: CTL EC $x_2(t)$

(**d**) IC4: CTL EO $x_2(t)$
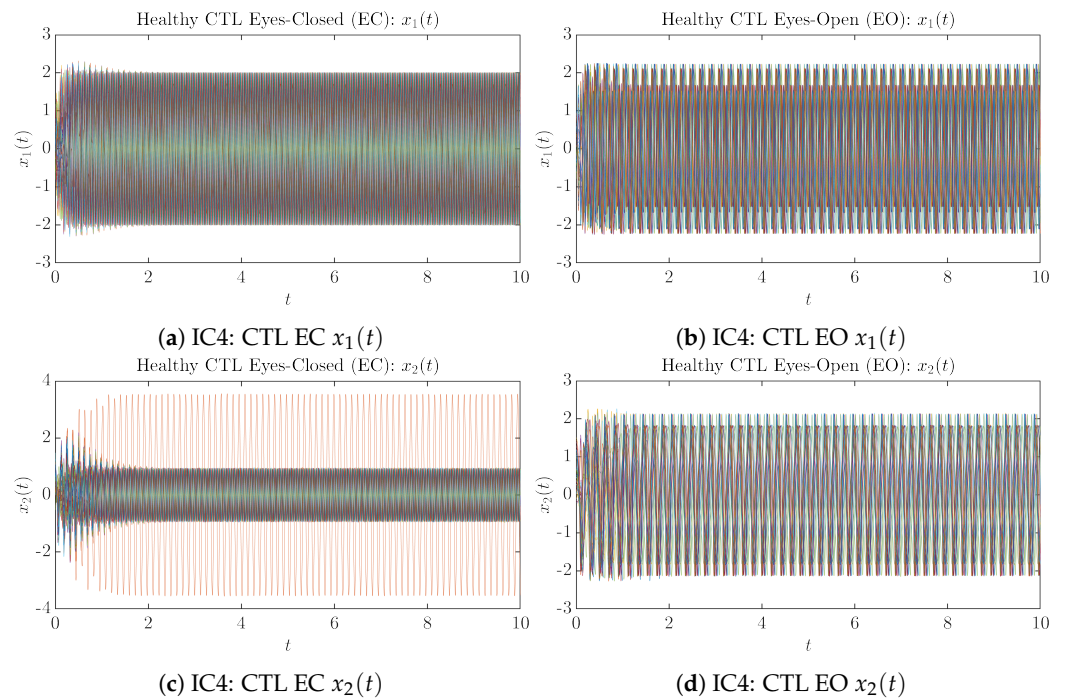
Figure A7. Initial Conditions No.4 (IC4): 50 sample trajectories of healthy CTL subjects. Each single trajectory is labeled by a different color.

(**a**) IC4: AD EC $x_1(t)$

(**b**) IC4: AD EO $x_1(t)$

(**c**) IC4: AD EC $x_2(t)$

(**d**) IC4: AD EO $x_2(t)$
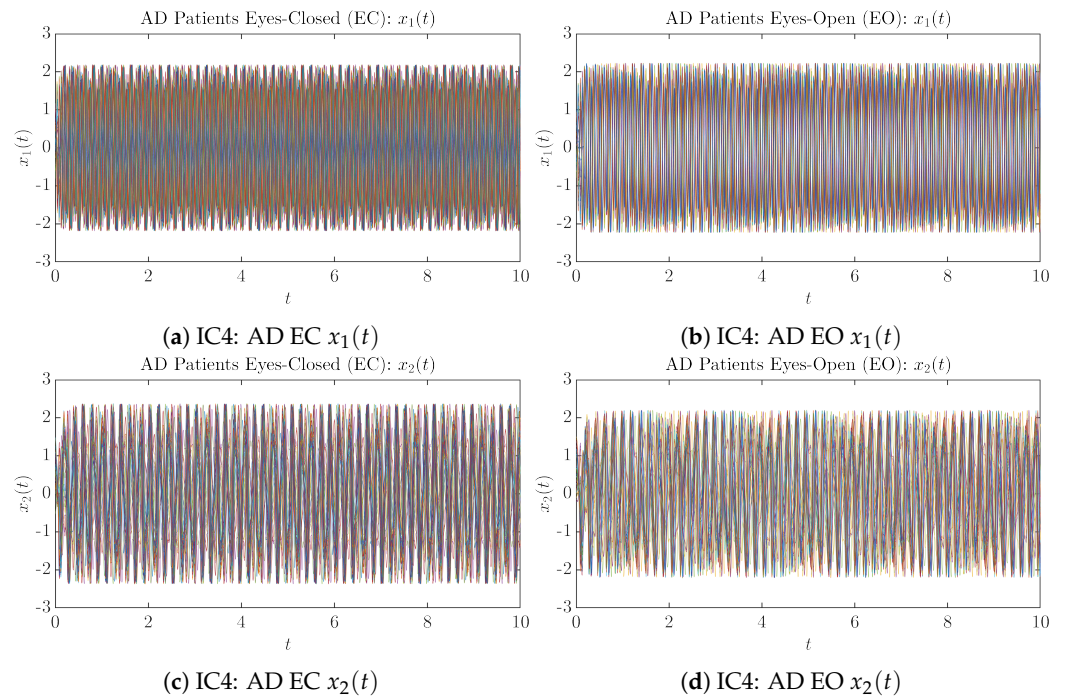
Figure A8. Initial Conditions No.4 (IC4): 50 sample trajectories of AD patients. Each single trajectory is labeled by a different color.

Appendix B.1.5. Initial Conditions No.5 (IC5)

(**a**) IC5: CTL EC $x_1(t)$

(**b**) IC5: CTL EO $x_1(t)$

(**c**) IC5: CTL EC $x_2(t)$

(**d**) IC5: CTL EO $x_2(t)$
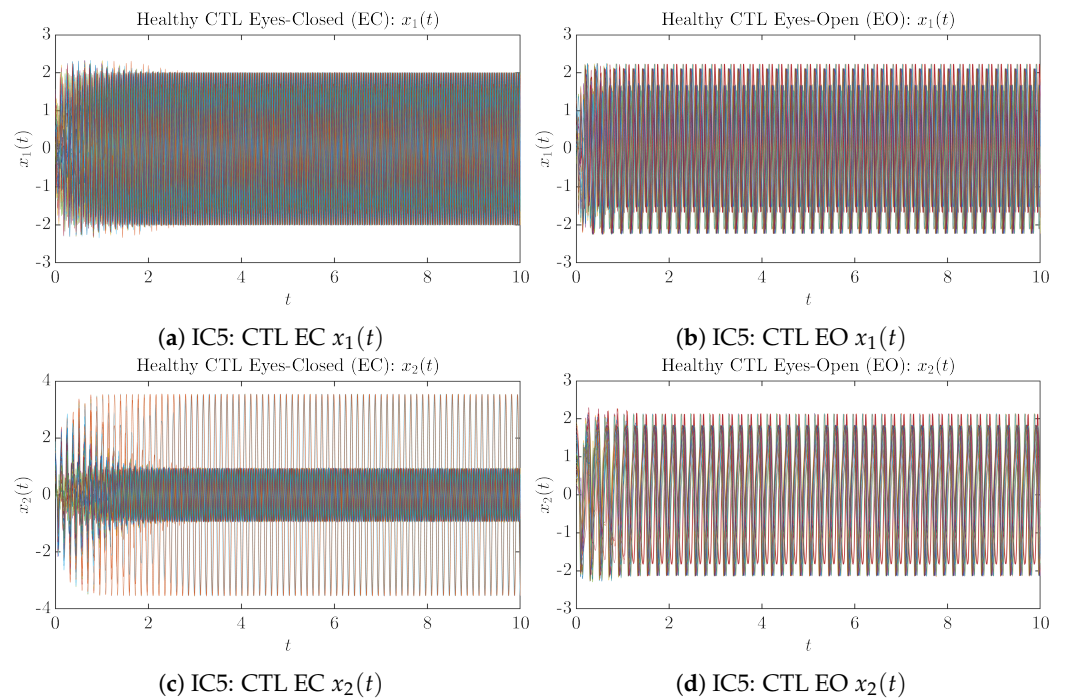
Figure A9. Initial Conditions No.5 (IC5): 50 sample trajectories of healthy CTL subjects. Each single trajectory is labeled by a different color.

(**a**) IC5: AD EC $x_1(t)$

(**b**) IC5: AD EO $x_1(t)$

(**c**) IC5: AD EC $x_2(t)$

(**d**) IC5: AD EO $x_2(t)$
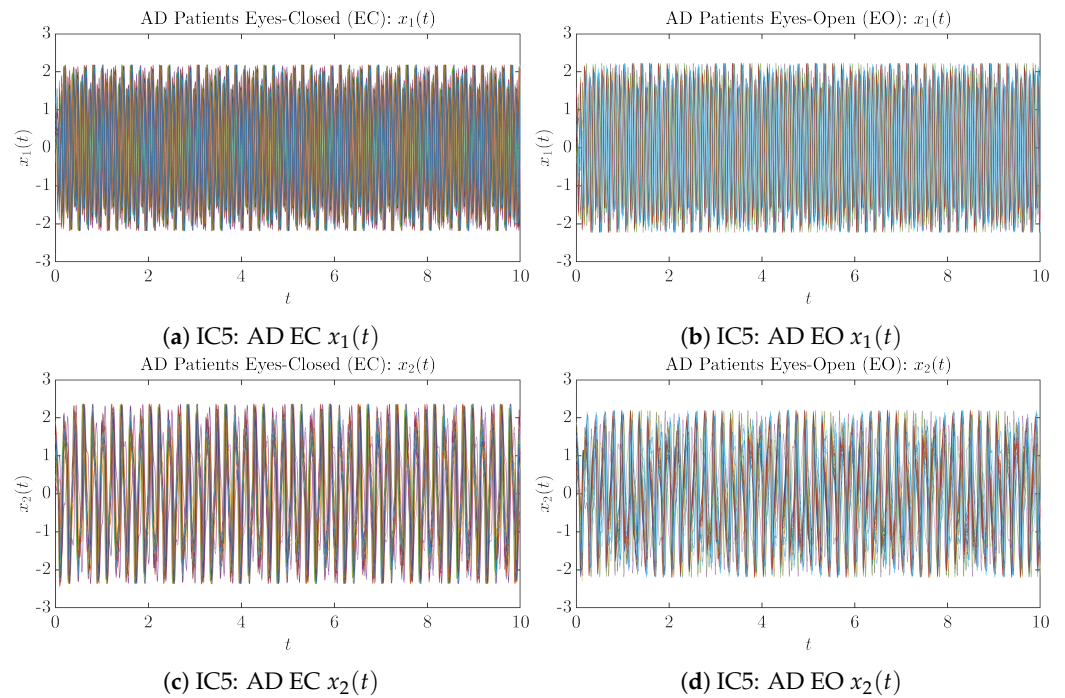
Figure A10. Initial Conditions No.5 (IC5): 50 sample trajectories of AD patients. Each single trajectory is labeled by a different color.

Appendix B.1.6. Initial Conditions No.6 (IC6)

(**a**) IC6: CTL EC $x_1(t)$

(**b**) IC6: CTL EO $x_1(t)$

(**c**) IC6: CTL EC $x_2(t)$

(**d**) IC6: CTL EO $x_2(t)$

Figure A11. Initial Conditions No.6 (IC6): 50 sample trajectories of healthy CTL subjects. Each single trajectory is labeled by a different color.

(**a**) IC6: AD EC $x_1(t)$

(**b**) IC6: AD EO $x_1(t)$

(**c**) IC6: AD EC $x_2(t)$

(**d**) IC6: AD EO $x_2(t)$

Figure A12. Initial Conditions No.6 (IC6): 50 sample trajectories of AD patients. Each single trajectory is labeled by a different color.

Appendix B.2. Time Evolution of PDF $p(x_1, t)$ and $p(x_2, t)$

Appendix B.2.1. Initial Conditions No.1 (IC1)

(**a**) IC1: CTL EC $p(x_1, t)$

(**b**) IC1: CTL EO $p(x_1, t)$

(**c**) IC1: CTL EC $p(x_2, t)$

(**d**) IC1: CTL EO $p(x_2, t)$

Figure A13. Initial Conditions No.1 (IC1): Time evolution of estimated PDFs of healthy CTL subjects.

(**a**) IC1: AD EC $p(x_1, t)$

(**b**) IC1: AD EO $p(x_1, t)$

(**c**) IC1: AD EC $p(x_2, t)$

(**d**) IC1: AD EO $p(x_2, t)$
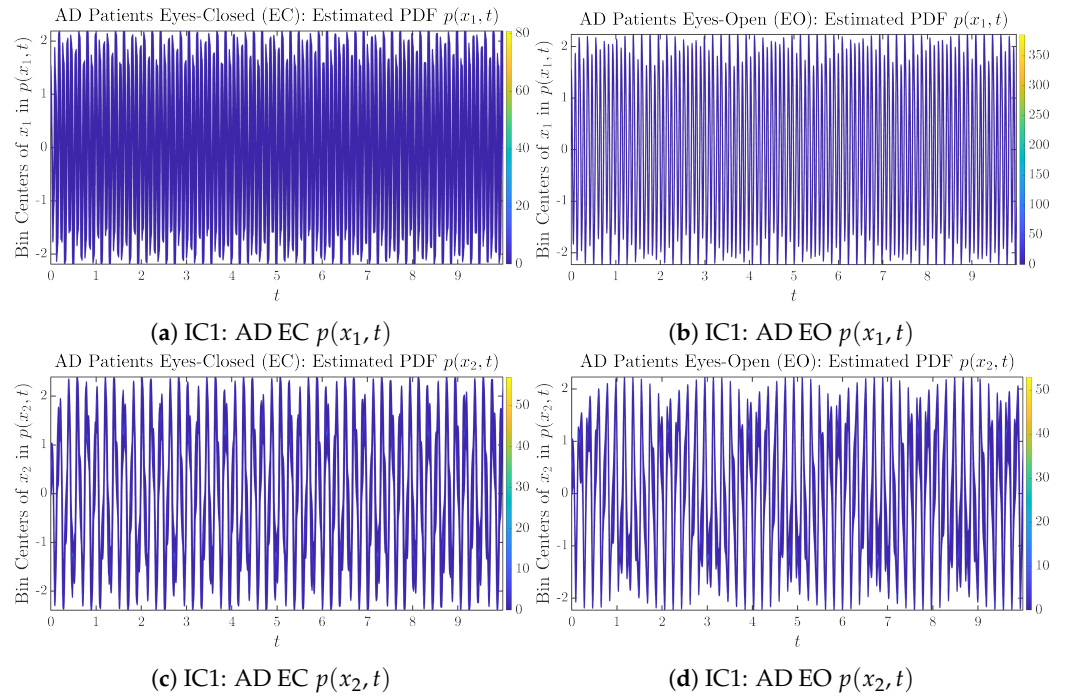
Figure A14. Initial Conditions No.1 (IC1): Time evolution of estimated PDFs of AD patients.

Appendix B.2.2. Initial Conditions No.2 (IC2)

(**a**) IC2: CTL EC $p(x_1, t)$

(**b**) IC2: CTL EO $p(x_1, t)$

(**c**) IC2: CTL EC $p(x_2, t)$

(**d**) IC2: CTL EO $p(x_2, t)$
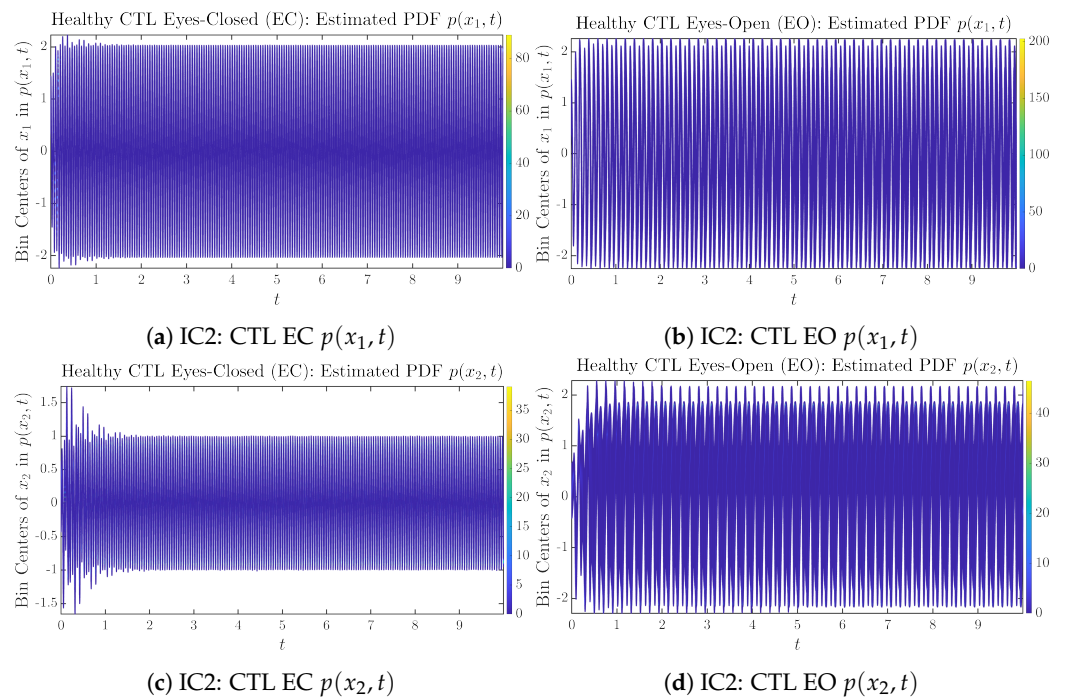
Figure A15. Initial Conditions No.2 (IC2): Time evolution of estimated PDFs of healthy CTL subjects.

(**a**) IC2: AD EC $p(x_1, t)$

(**b**) IC2: AD EO $p(x_1, t)$

(**c**) IC2: AD EC $p(x_2, t)$

(**d**) IC2: AD EO $p(x_2, t)$
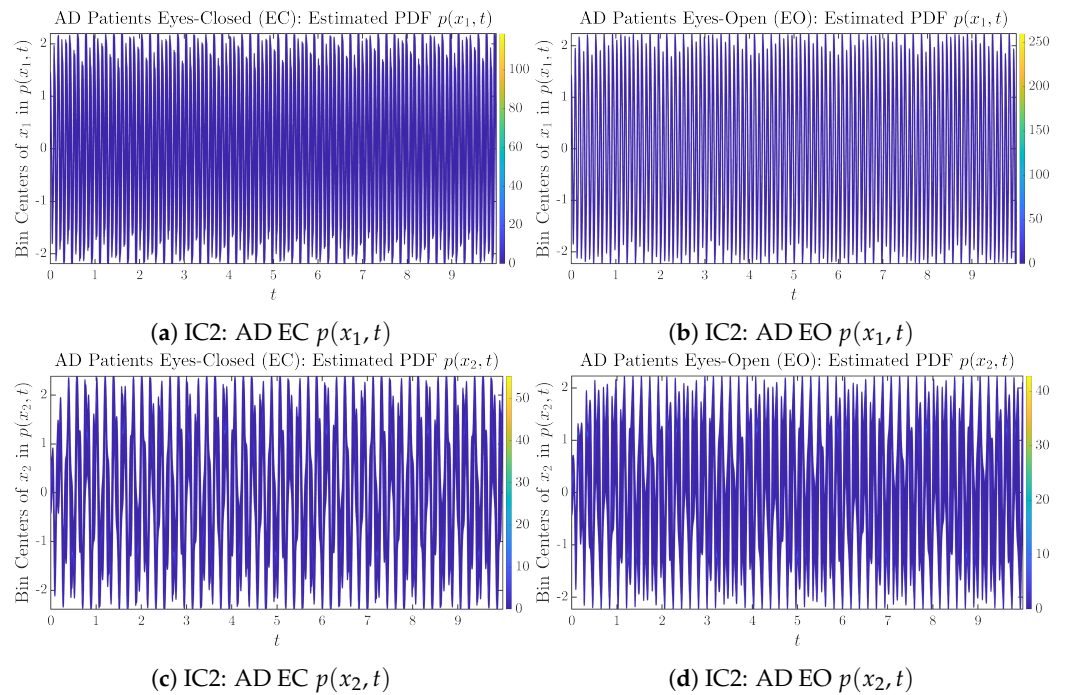
Figure A16. Initial Conditions No.2 (IC2): Time evolution of estimated PDFs of AD patients.

Appendix B.2.3. Initial Conditions No.3 (IC3)

(**a**) IC3: CTL EC $p(x_1, t)$

(**b**) IC3: CTL EO $p(x_1, t)$

(**c**) IC3: CTL EC $p(x_2, t)$

(**d**) IC3: CTL EO $p(x_2, t)$
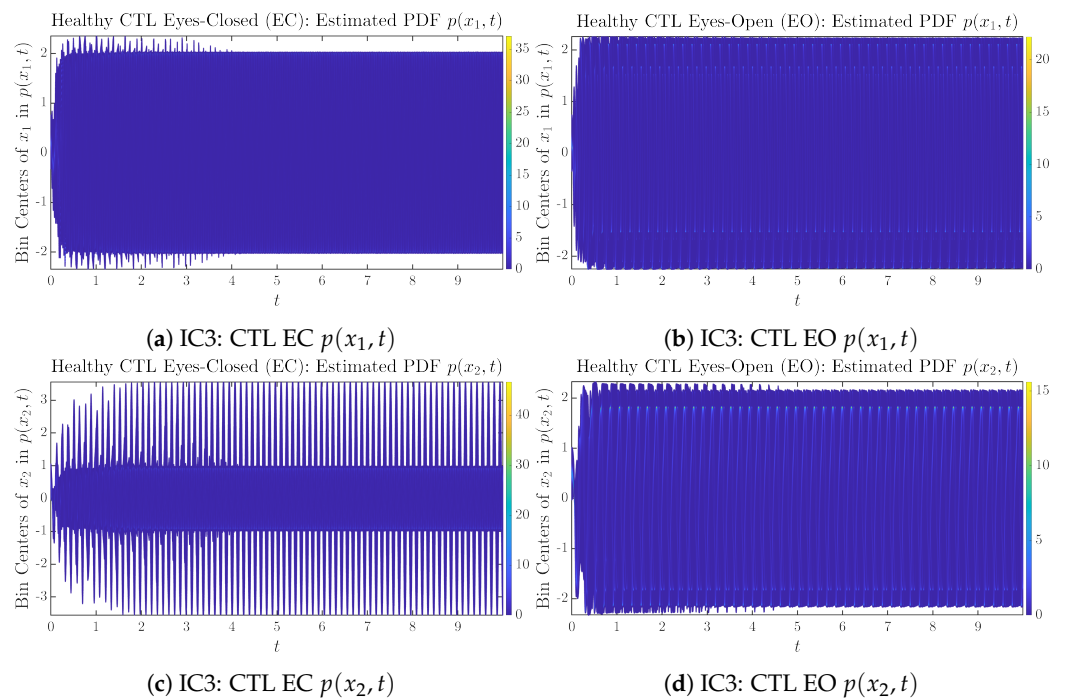
Figure A17. Initial Conditions No.3 (IC3): Time evolution of estimated PDFs of healthy CTL subjects.

(**a**) IC3: AD EC $p(x_1, t)$

(**b**) IC3: AD EO $p(x_1, t)$

(**c**) IC3: AD EC $p(x_2, t)$

(**d**) IC3: AD EO $p(x_2, t)$
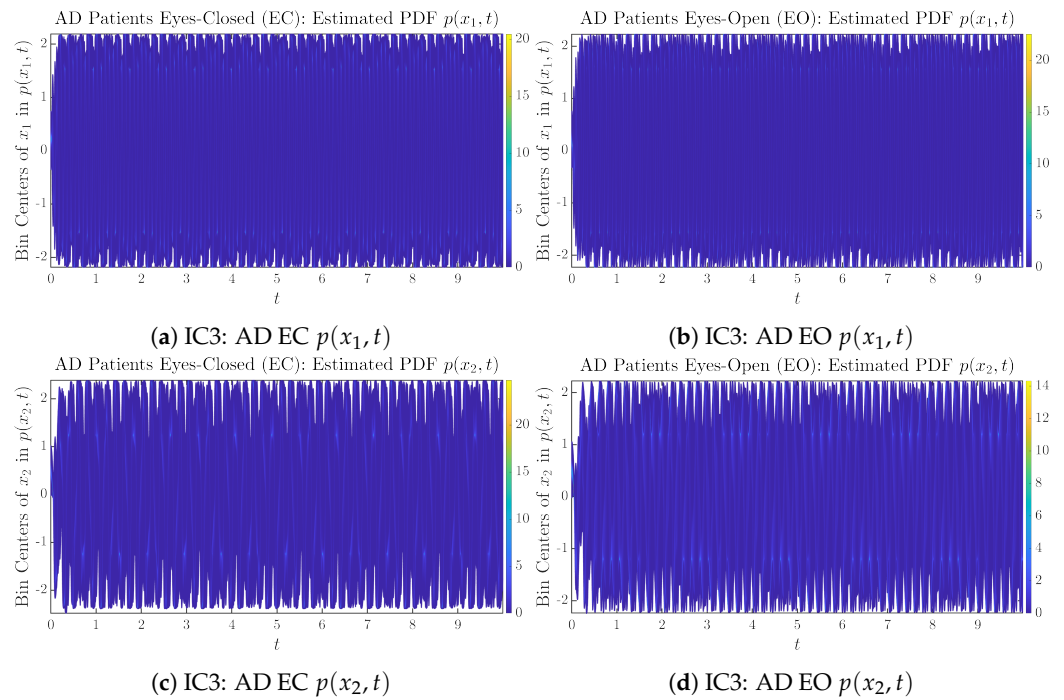
Figure A18. Initial Conditions No.3 (IC3): Time evolution of estimated PDFs of AD patients.

Appendix B.2.4. Initial Conditions No.4 (IC4)

(**a**) IC4: CTL EC $p(x_1, t)$

(**b**) IC4: CTL EO $p(x_1, t)$

(**c**) IC4: CTL EC $p(x_2, t)$

(**d**) IC4: CTL EO $p(x_2, t)$
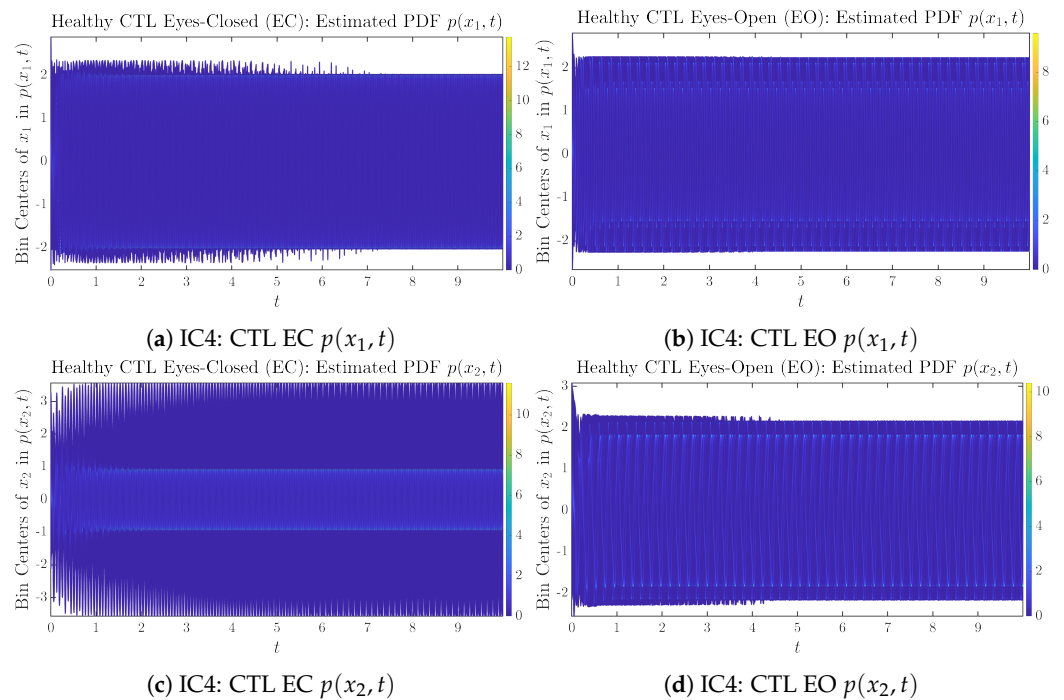
Figure A19. Initial Conditions No.4 (IC4): Time evolution of estimated PDFs of healthy CTL subjects.

(**a**) IC4: AD EC $p(x_1, t)$

(**b**) IC4: AD EO $p(x_1, t)$

(**c**) IC4: AD EC $p(x_2, t)$

(**d**) IC4: AD EO $p(x_2, t)$
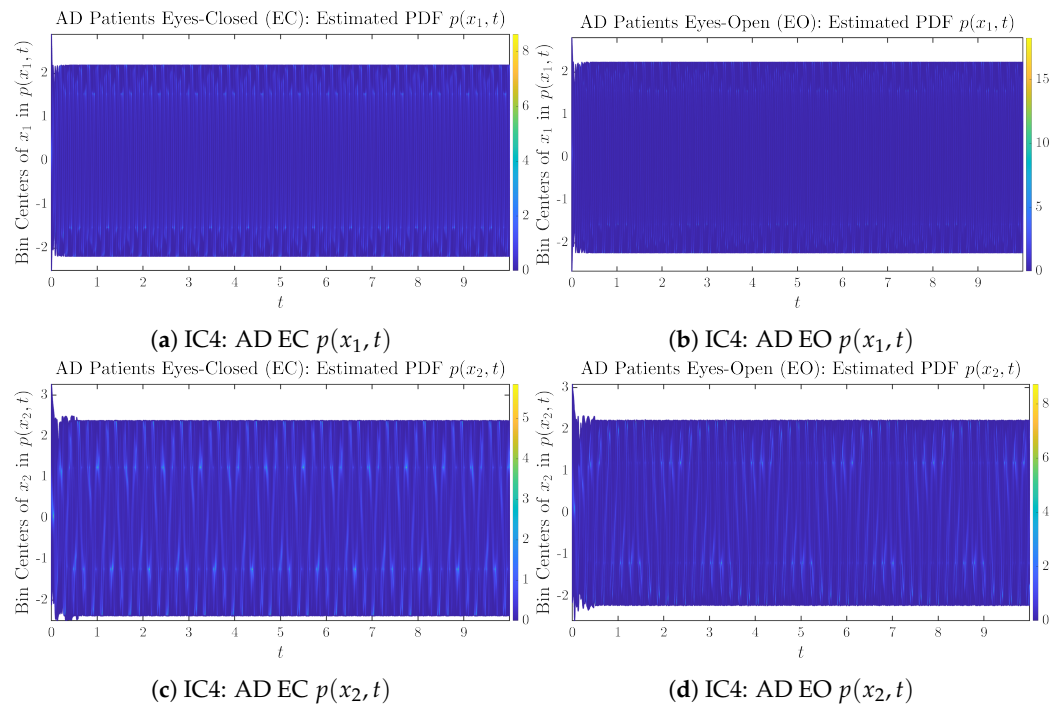
Figure A20. Initial Conditions No.4 (IC4): Time evolution of estimated PDFs of AD patients.

Appendix B.2.5. Initial Conditions No.5 (IC5)

(**a**) IC5: CTL EC $p(x_1, t)$

(**b**) IC5: CTL EO $p(x_1, t)$

(**c**) IC5: CTL EC $p(x_2, t)$

(**d**) IC5: CTL EO $p(x_2, t)$
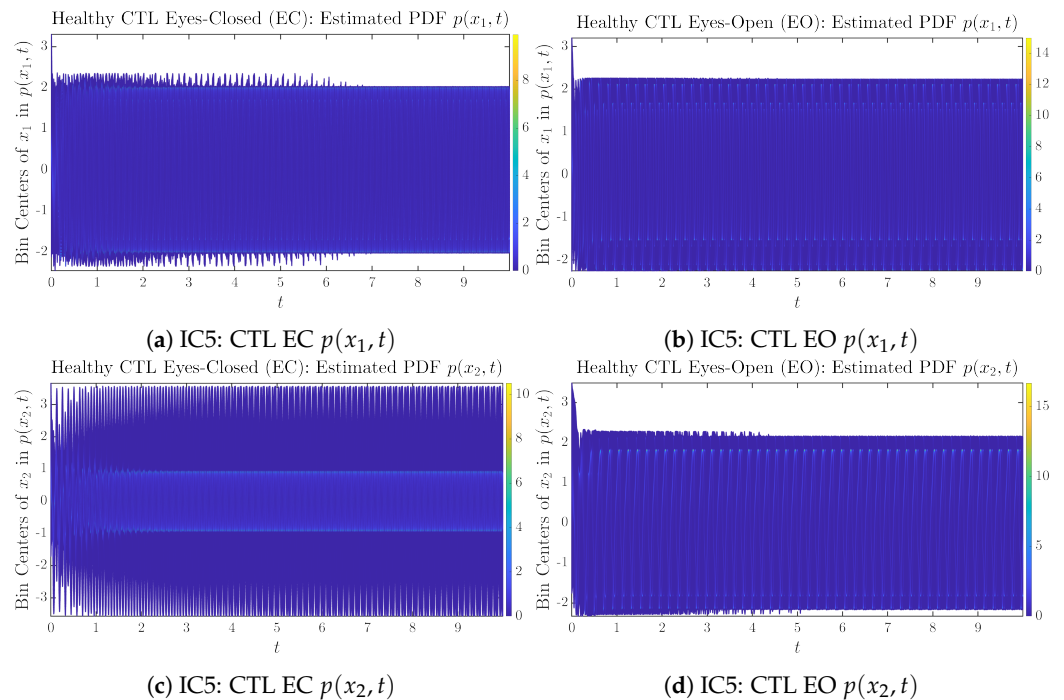
Figure A21. Initial Conditions No.5 (IC5): Time evolution of estimated PDFs of healthy CTL subjects.

(**a**) IC5: AD EC $p(x_1, t)$

(**b**) IC5: AD EO $p(x_1, t)$

(**c**) IC5: AD EC $p(x_2, t)$

(**d**) IC5: AD EO $p(x_2, t)$
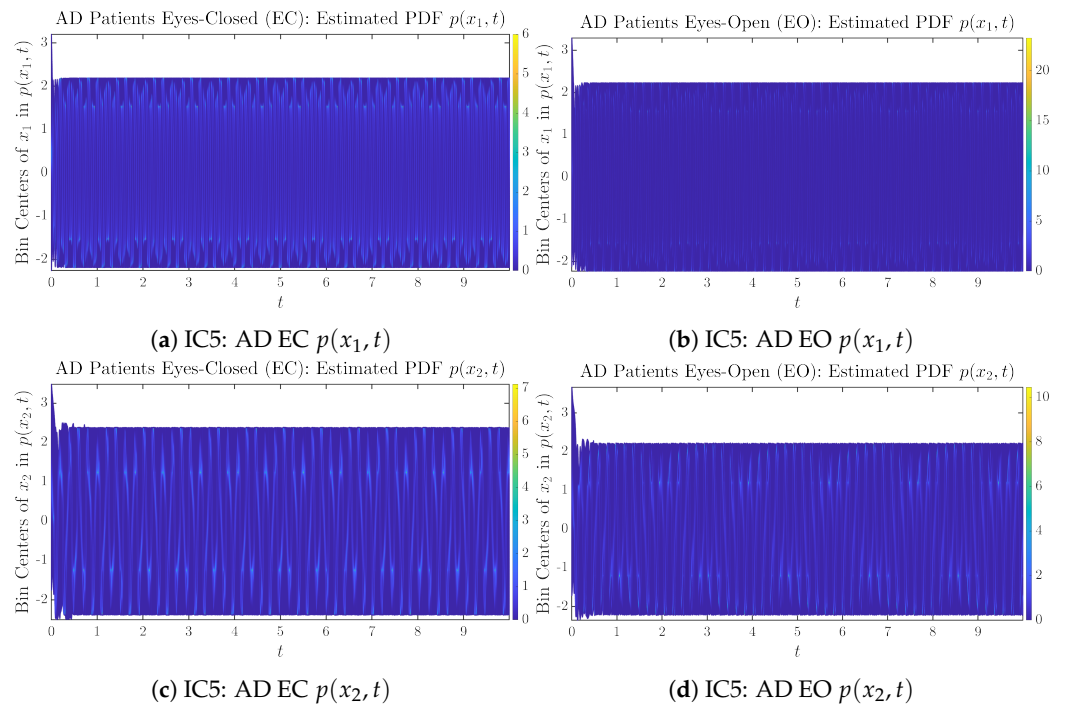
Figure A22. Initial Conditions No.5 (IC5): Time evolution of estimated PDFs of AD patients.

Appendix B.2.6. Initial Conditions No.6 (IC6)

(**a**) IC6: CTL EC $p(x_1, t)$

(**b**) IC6: CTL EO $p(x_1, t)$

(**c**) IC6: CTL EC $p(x_2, t)$

(**d**) IC6: CTL EO $p(x_2, t)$
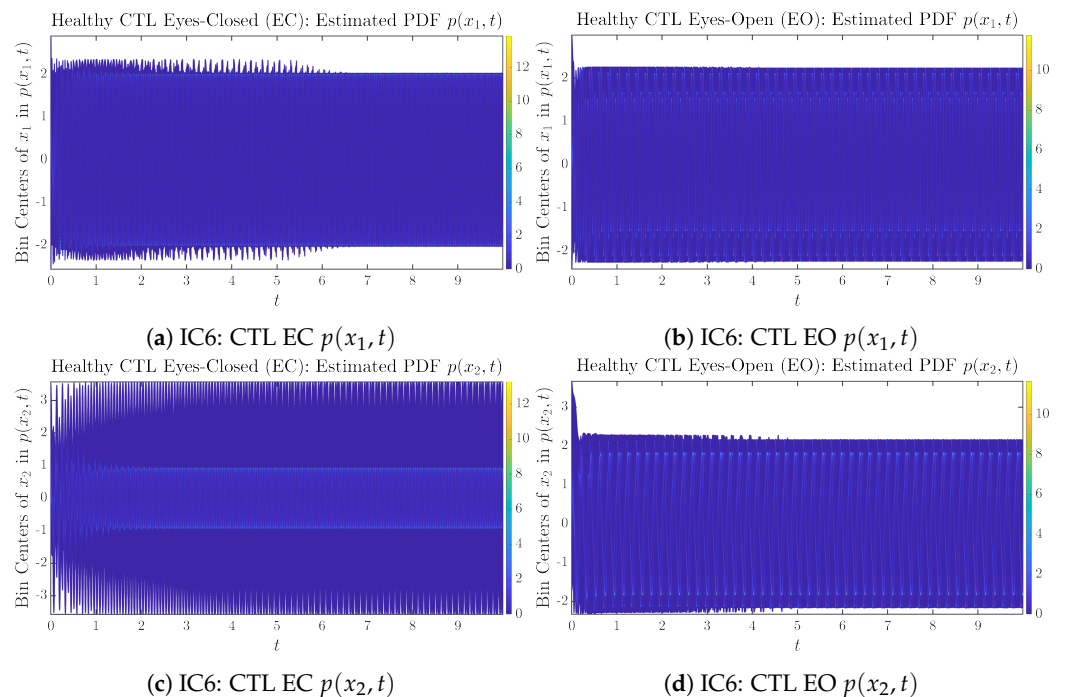
Figure A23. Initial Conditions No.6 (IC6): Time evolution of estimated PDFs of healthy CTL subjects.

(**a**) IC6: AD EC $p(x_1, t)$

(**b**) IC6: AD EO $p(x_1, t)$

(**c**) IC6: AD EC $p(x_2, t)$

(**d**) IC6: AD EO $p(x_2, t)$
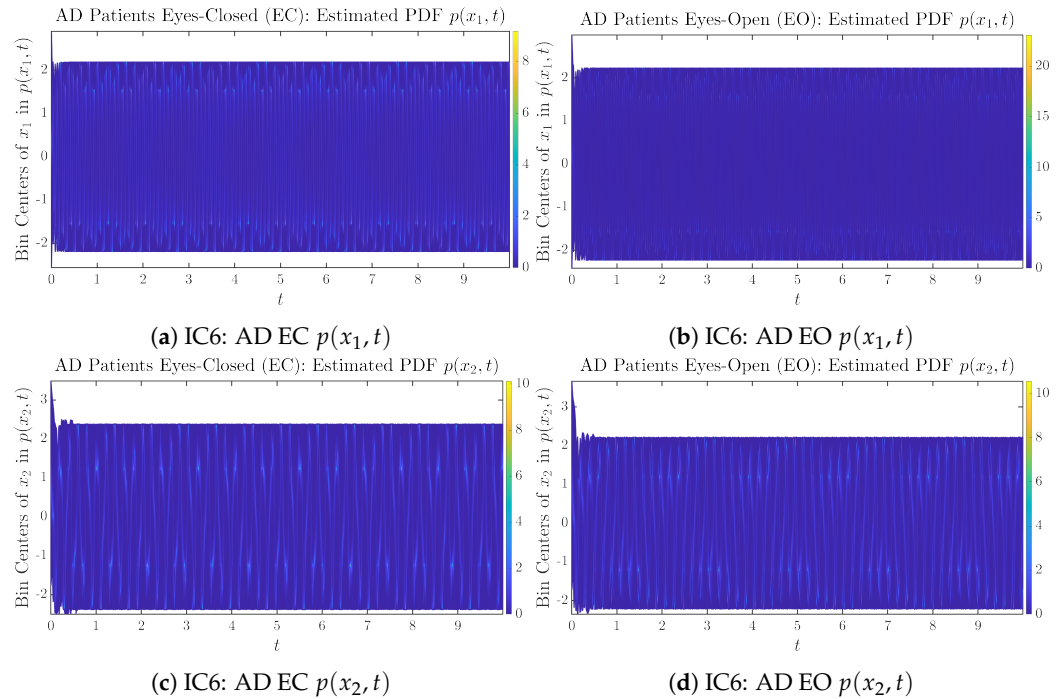
Figure A24. Initial Conditions No.6 (IC6): Time evolution of estimated PDFs of AD patients.

Appendix B.3. Information Rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$

Appendix B.3.1. Time Evolution: Information Rates

Initial Conditions No.1 (IC1)

(**a**) IC1: health CTL subjects.

(**b**) IC1: AD patients.
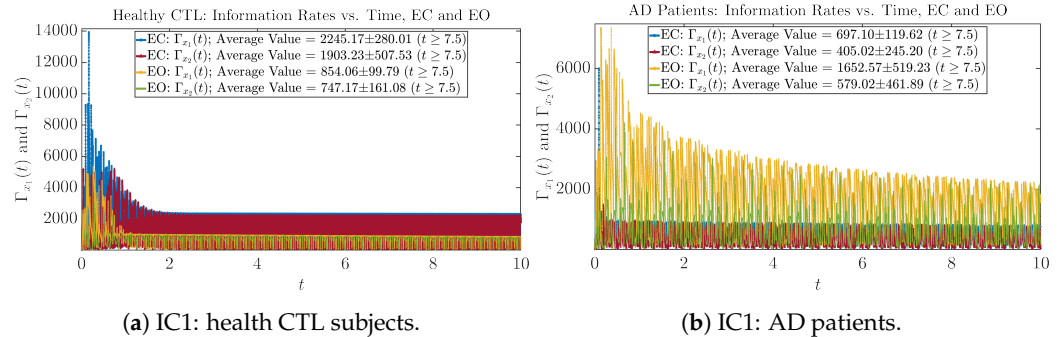
Figure A25. Initial Conditions No.1 (IC1): Information rates along time of CTL and AD subjects.

Initial Conditions No.2 (IC2)

(**a**) IC2: health CTL subjects.
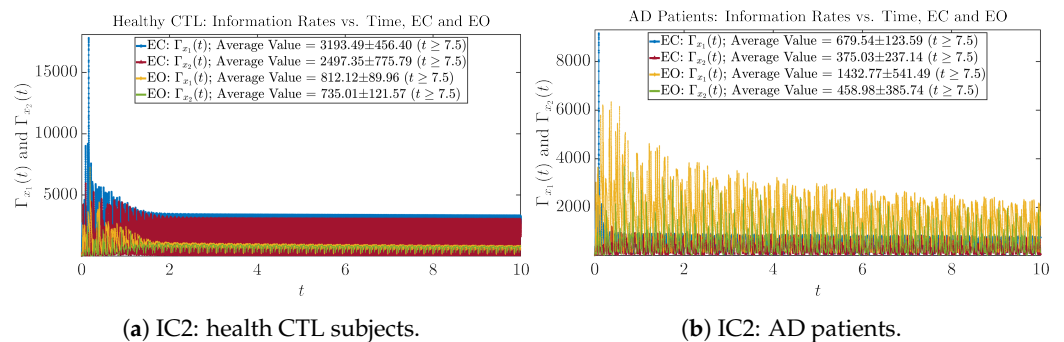
(**b**) IC2: AD patients.

Figure A26. Initial Conditions No.2 (IC2): Information rates along time of CTL and AD subjects.

Initial Conditions No.3 (IC3)
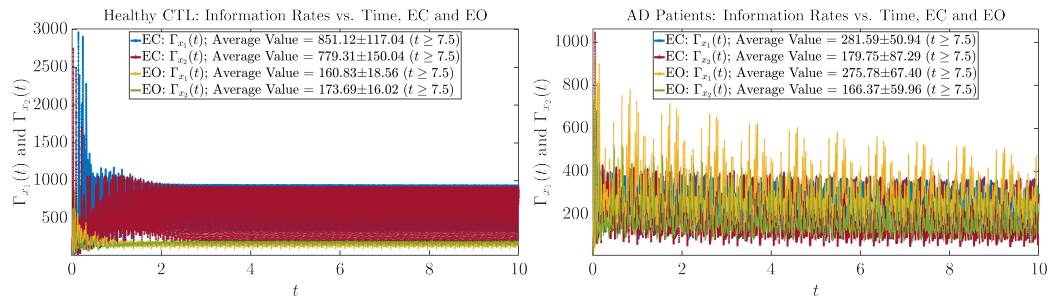
(**a**) IC3: health CTL subjects.

(**b**) IC3: AD patients.

Figure A27. Initial Conditions No.3 (IC3): Information rates along time of CTL and AD subjects.

Initial Conditions No.4 (IC4)
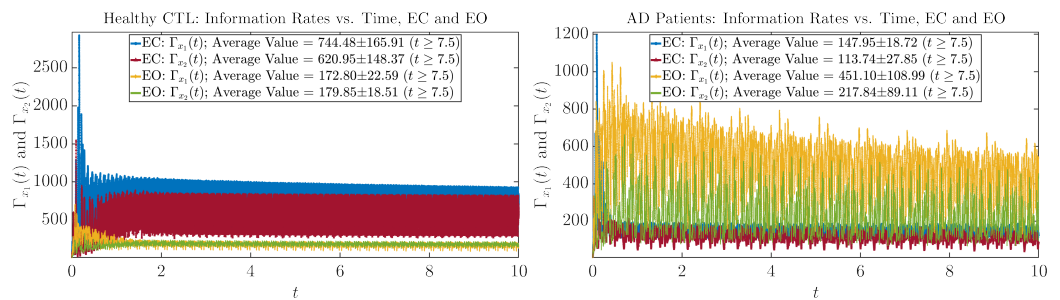
(**a**) IC4: health CTL subjects.

(**b**) IC4: AD patients.

Figure A28. Initial Conditions No.4 (IC4): Information rates along time of CTL and AD subjects.

Initial Conditions No.5 (IC5)
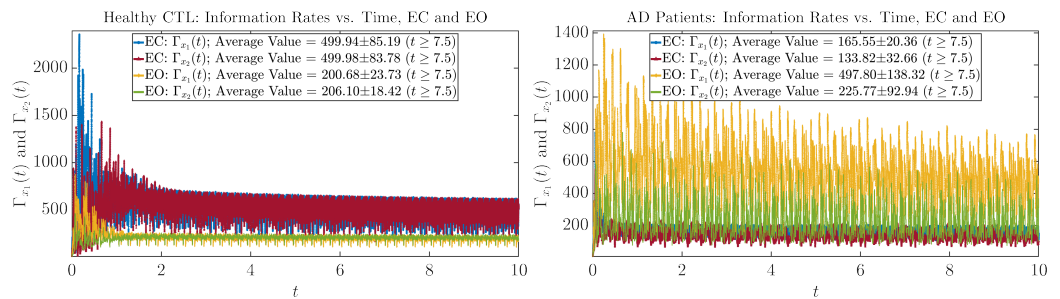
(**a**) IC5: health CTL subjects.

(**b**) IC5: AD patients.

Figure A29. Initial Conditions No.5 (IC5): Information rates along time of CTL and AD subjects.

Initial Conditions No.6 (IC6)
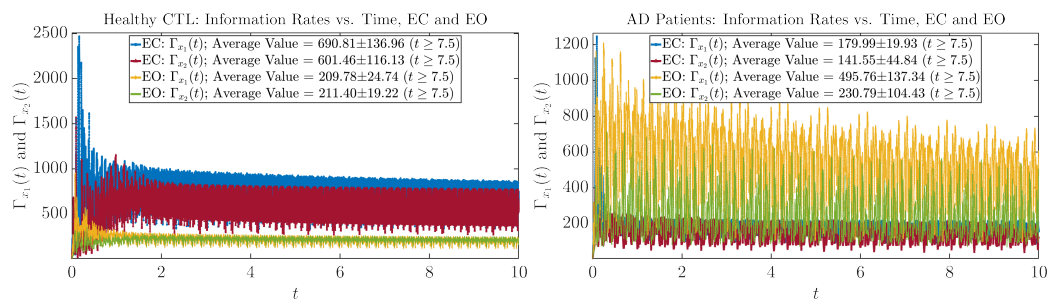
(**a**) IC6: health CTL subjects.

(**b**) IC6: AD patients.

Figure A30. Initial Conditions No.6 (IC6): Information rates along time of CTL and AD subjects.

Appendix B.3.2. Empirical Probability Distribution: Information Rates (for $t \geq 7.5$)

(**a**) Initial Conditions No.1 (IC1)

(**b**) Initial Conditions No.2 (IC2)

(**c**) Initial Conditions No.3 (IC3)

(**d**) Initial Conditions No.4 (IC4)

(**e**) Initial Conditions No.5 (IC5)

(**f**) Initial Conditions No.6 (IC6)
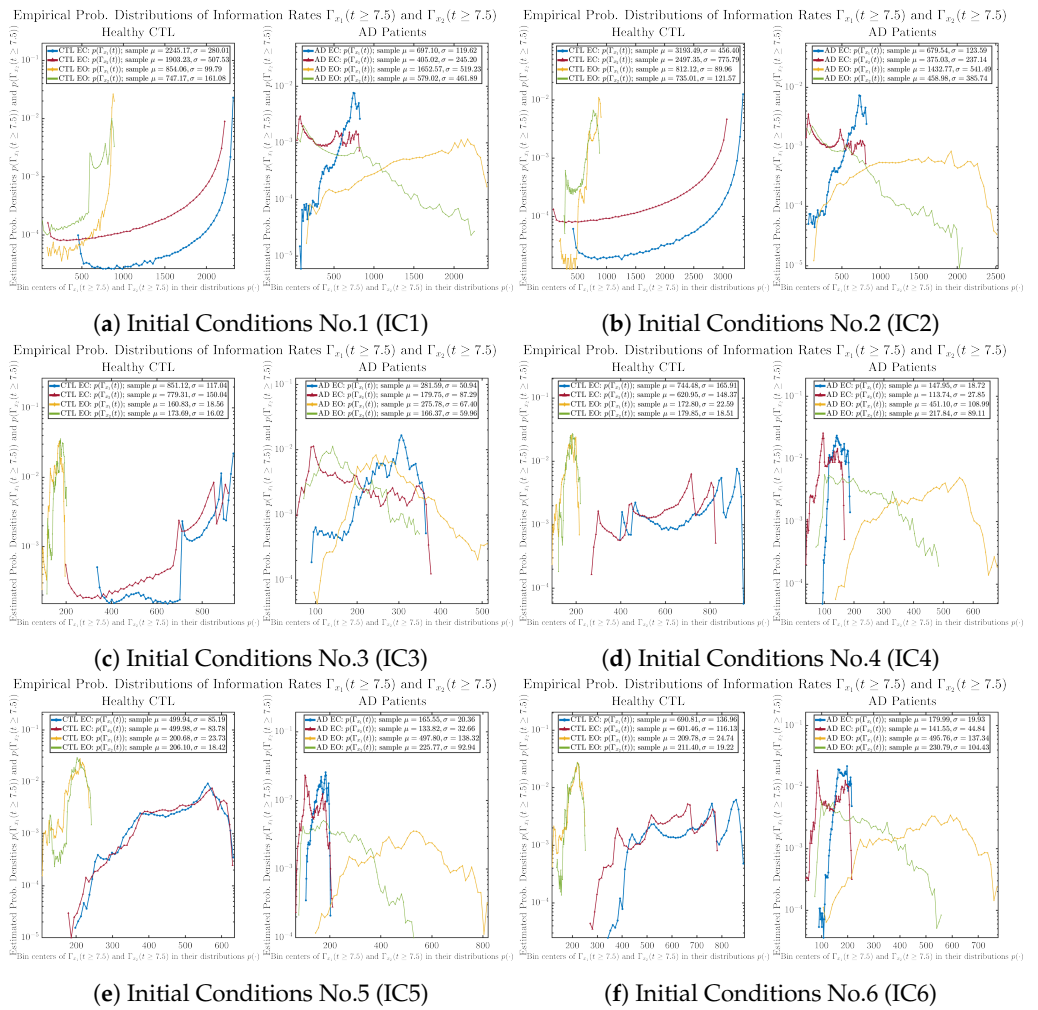
Figure A31. Empirical probability distributions of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$).

Appendix B.3.3. Phase Portraits: Information Rates (for $t \geq 7.5$)

(**a**) Initial Conditions No.1 (IC1)

(**b**) Initial Conditions No.2 (IC2)

(**c**) Initial Conditions No.3 (IC3)

(**d**) Initial Conditions No.4 (IC4)

(**e**) Initial Conditions No.5 (IC5)

(**f**) Initial Conditions No.6 (IC6)
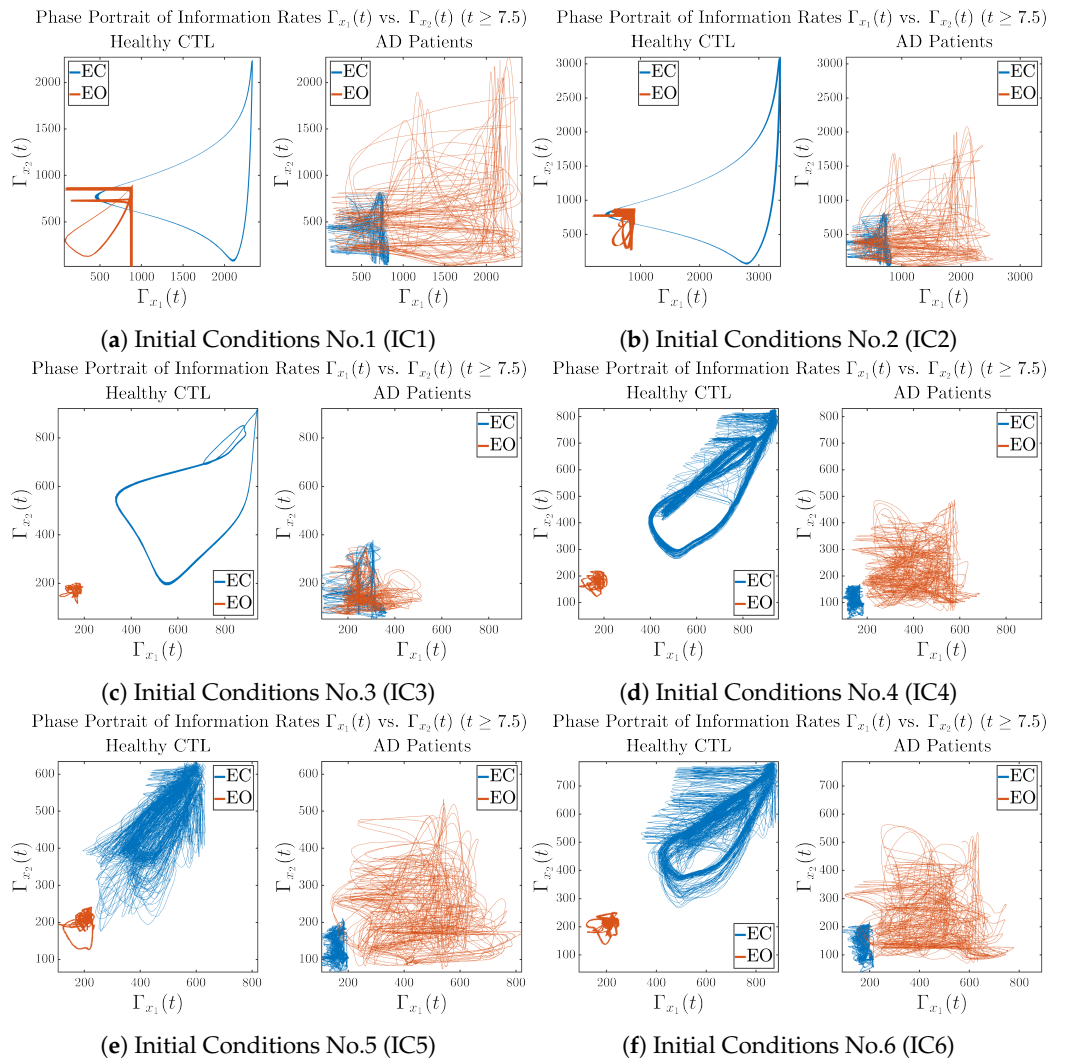
Figure A32. Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$).

Initial Conditions No.1 (IC1):

(**a**) IC1: health CTL subjects.

(**b**) IC1: AD patients.
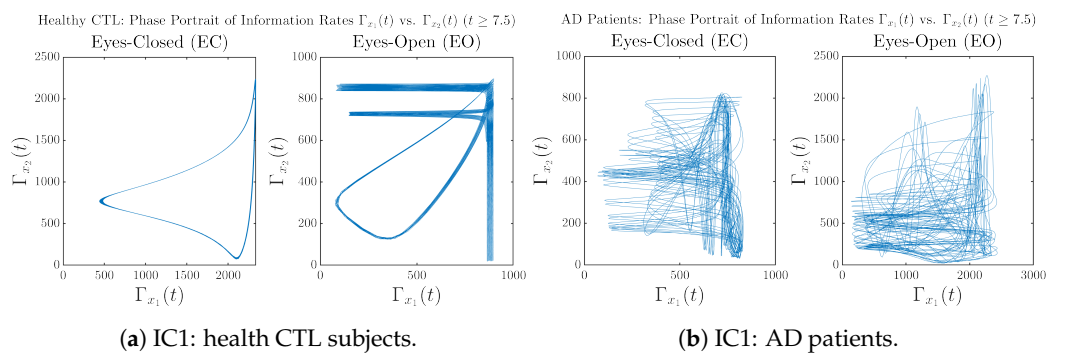
Figure A33. Initial Conditions No.1 (IC1): Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.2 (IC2):
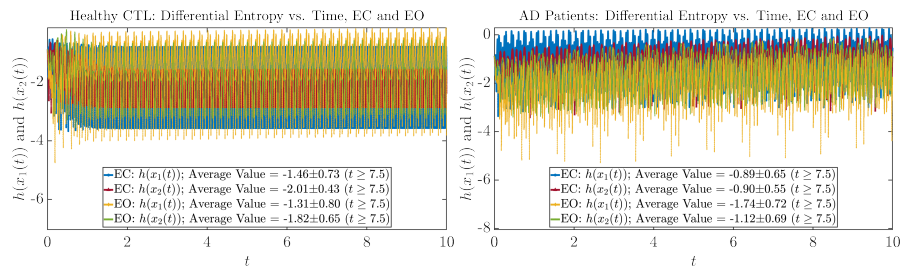

(**a**) IC2: health CTL subjects.

(**b**) IC2: AD patients.
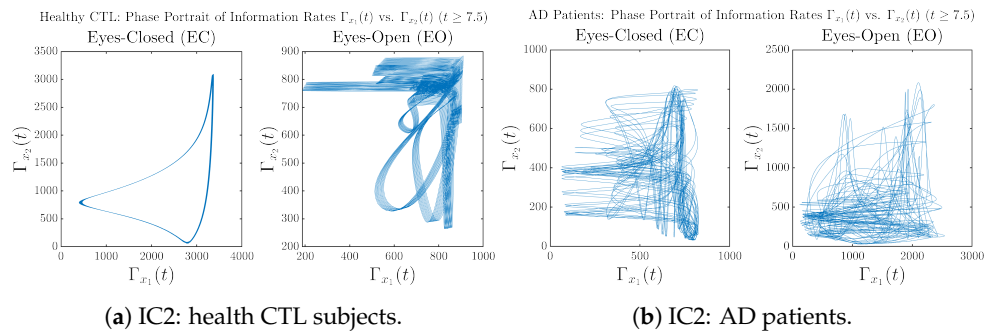
Figure A34. Initial Conditions No.2 (IC2): Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.3 (IC3):

(**a**) IC3: health CTL subjects.

(**b**) IC3: AD patients.
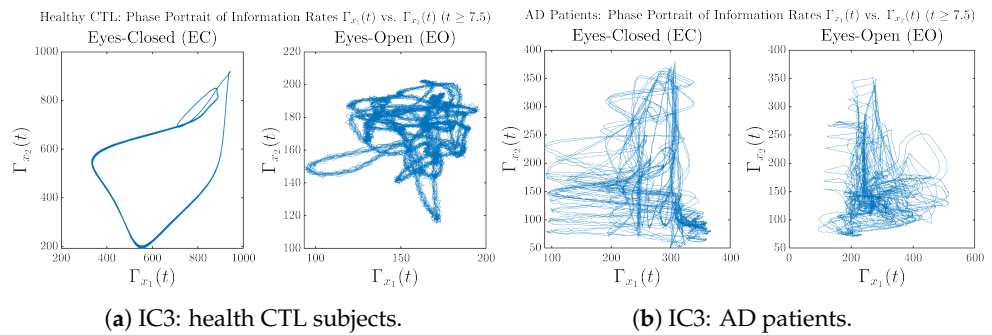
Figure A35. Initial Conditions No.3 (IC3): Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.4 (IC4):

(**a**) IC4: health CTL subjects.
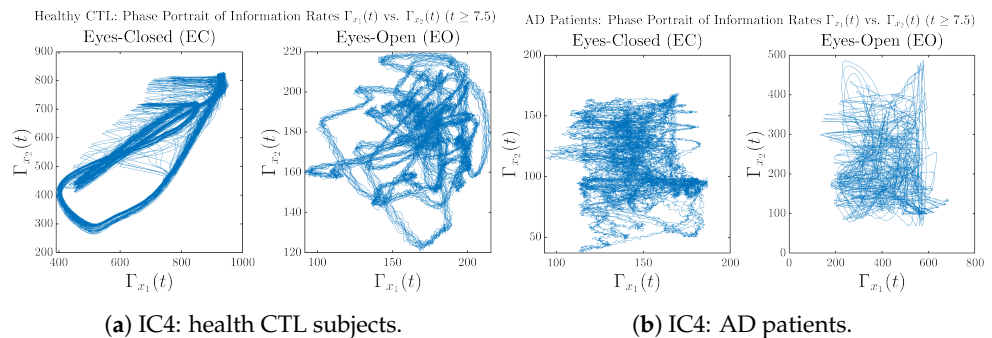
(**b**) IC4: AD patients.

Figure A36. Initial Conditions No.4 (IC4): Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.5 (IC5):
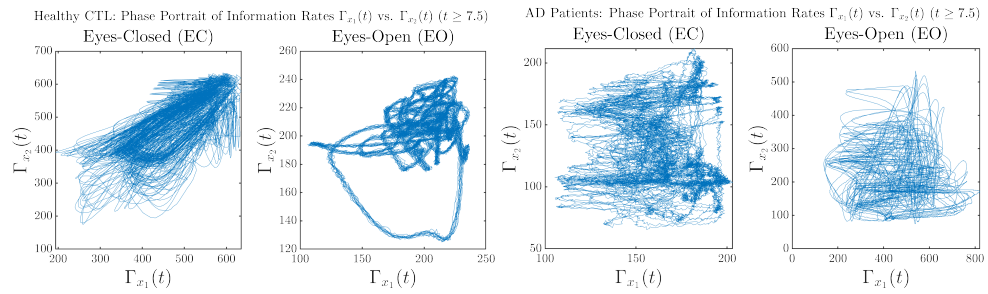
(**a**) IC5: health CTL subjects.   (**b**) IC5: AD patients.

Figure A37. Initial Conditions No.5 (IC5): Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.6 (IC6):
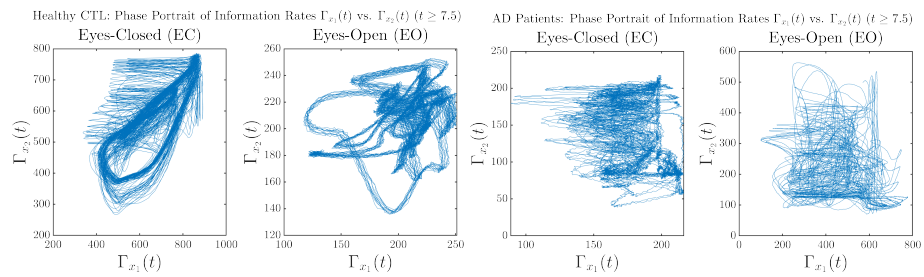
(**a**) IC6: health CTL subjects.   (**b**) IC6: AD patients.

Figure A38. Initial Conditions No.6 (IC6): Phase portraits of information rates $\Gamma_{x_1}(t)$ vs. $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Appendix B.3.4. Power Spectra: Information Rates (for $t \geq 7.5$)

Initial Conditions No.1 (IC1):
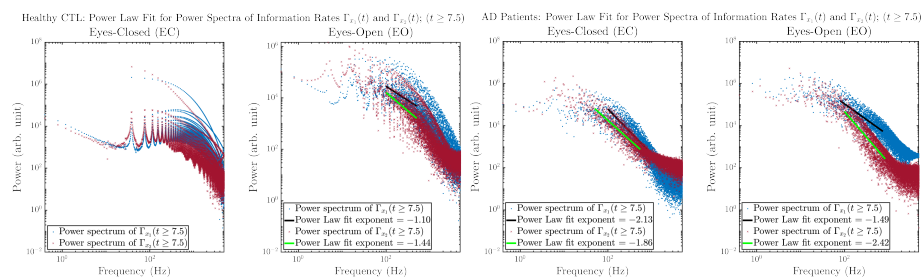
(**a**) IC1: health CTL subjects.   (**b**) IC1: AD patients.

Figure A39. Initial Conditions No.1 (IC1): Power law fit for power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.2 (IC2):
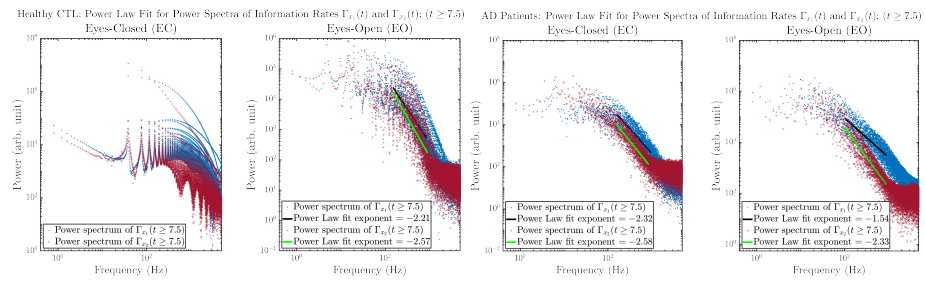
(**a**) IC2: health CTL subjects.   (**b**) IC2: AD patients.

Figure A40. Initial Conditions No.2 (IC2): Power law fit for power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.3 (IC3):
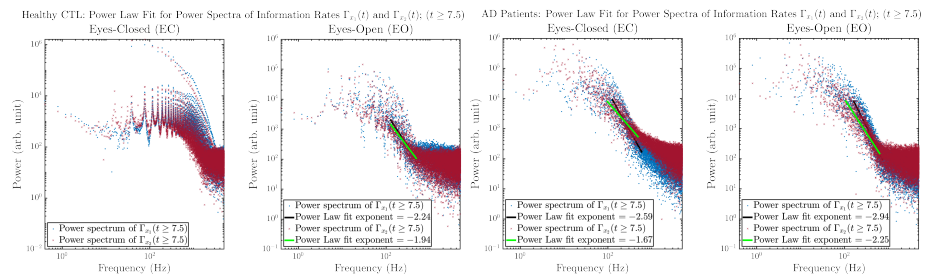
(**a**) IC3: health CTL subjects.   (**b**) IC3: AD patients.

Figure A41. Initial Conditions No.3 (IC3): Power law fit for power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.4 (IC4):
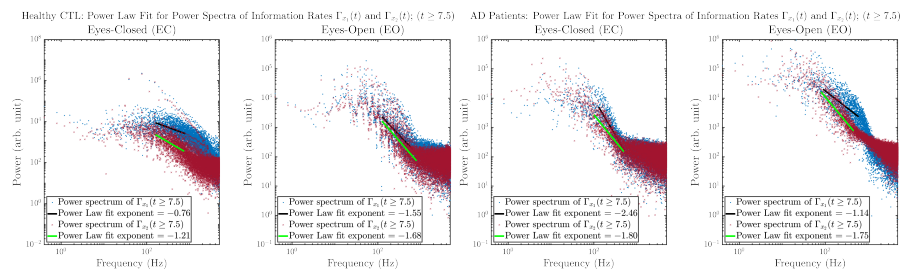
(**a**) IC4: health CTL subjects.   (**b**) IC4: AD patients.

Figure A42. Initial Conditions No.4 (IC4): Power law fit for power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.5 (IC5):
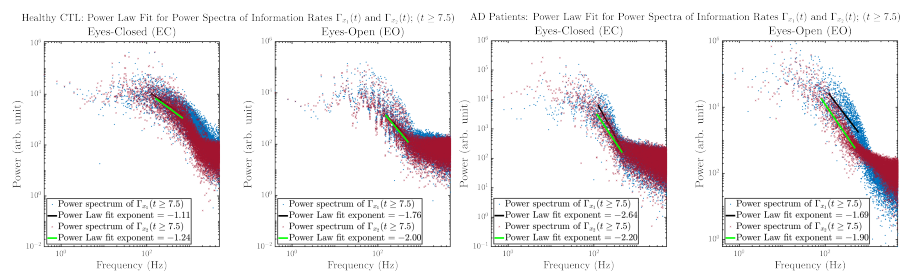
(**a**) IC5: health CTL subjects.   (**b**) IC5: AD patients.

Figure A43. Initial Conditions No.5 (IC5): Power law fit for power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.6 (IC6):
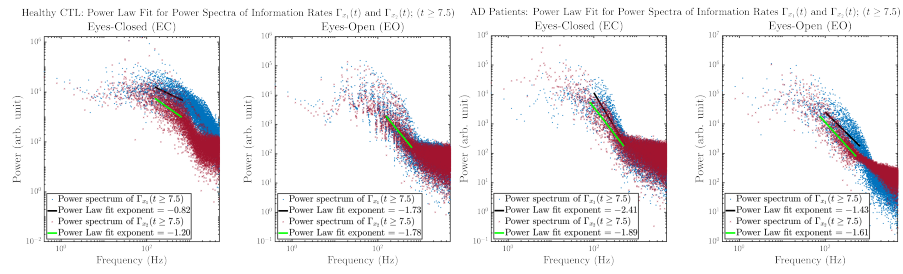
(**a**) IC6: health CTL subjects.　　　　　(**b**) IC6: AD patients.

Figure A44. Initial Conditions No.6 (IC6): Power law fit for power spectra of information rates $\Gamma_{x_1}(t)$ and $\Gamma_{x_2}(t)$ ($t \geq 7.5$) of CTL and AD subjects.

Appendix B.4. Shannon Differential Entropy of $p(x_1, t)$ and $p(x_2, t)$

Appendix B.4.1. Time Evolution: Shannon Differential Entropy

Initial Conditions No.1 (IC1):

(**a**) IC1: health CTL subjects.　　　　　(**b**) IC1: AD patients.

Figure A45. Initial Conditions No.1 (IC1): Shannon differential entropy along time of CTL and AD subjects.
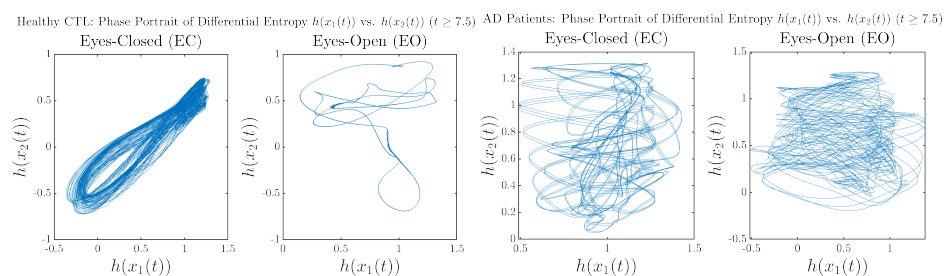
Initial Conditions No.2 (IC2):
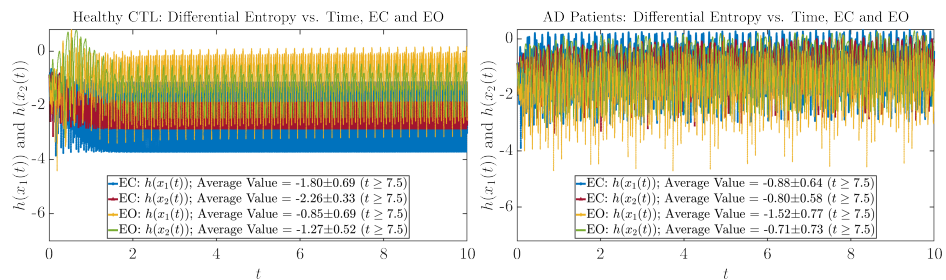
(**a**) IC2: health CTL subjects.　　　　　(**b**) IC2: AD patients.

Figure A46. Initial Conditions No.2 (IC2): Shannon differential entropy along time of CTL and AD subjects.

Initial Conditions No.3 (IC3):
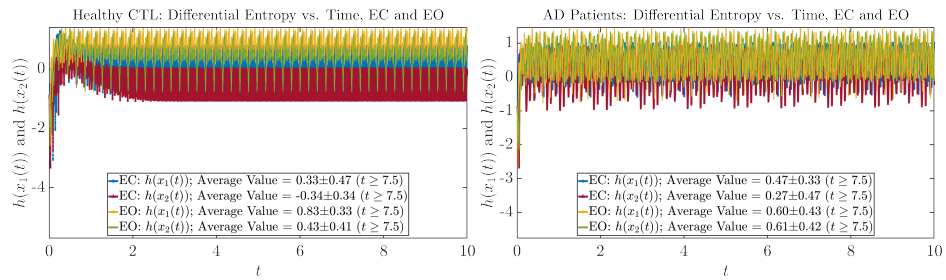
(**a**) IC3: health CTL subjects.

(**b**) IC3: AD patients.

Figure A47. Initial Conditions No.3 (IC3): Shannon differential entropy along time of CTL and AD subjects.
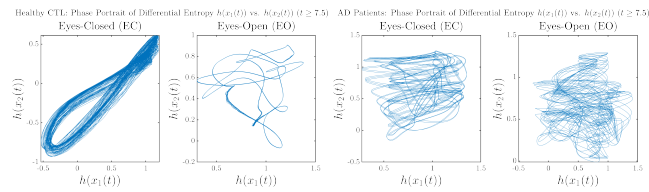
Initial Conditions No.4 (IC4):
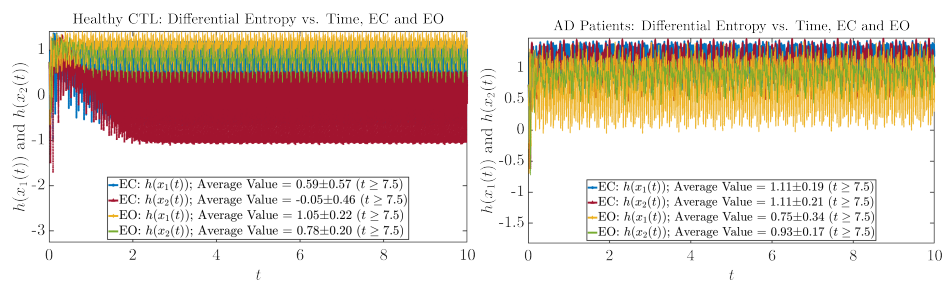
(**a**) IC4: health CTL subjects.

(**b**) IC4: AD patients.

Figure A48. Initial Conditions No.4 (IC4): Shannon differential entropy along time of CTL and AD subjects.

Initial Conditions No.5 (IC5):
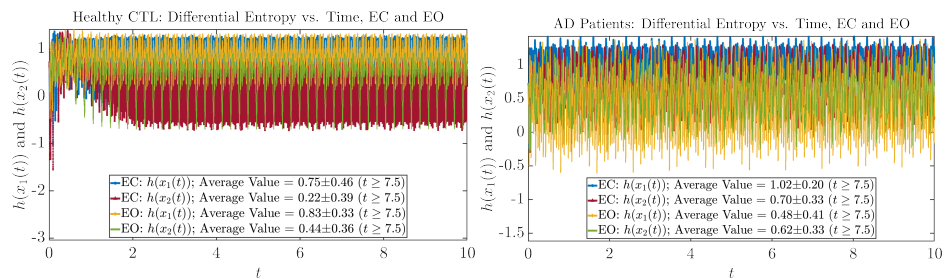
(**a**) IC5: health CTL subjects.

(**b**) IC5: AD patients.

Figure A49. Initial Conditions No.5 (IC5): Shannon differential entropy along time of CTL and AD subjects.
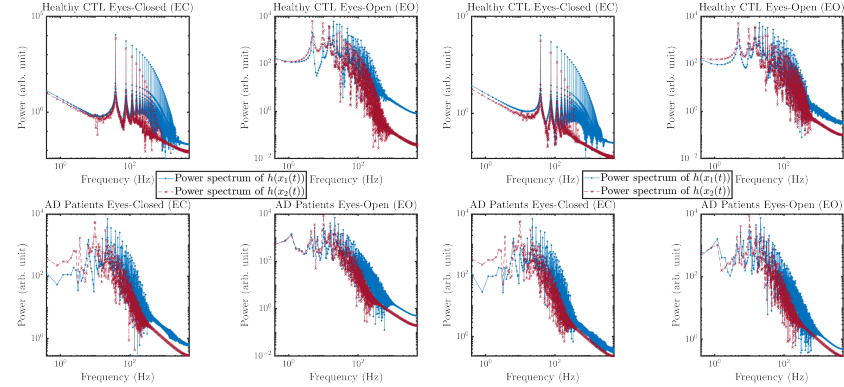
Initial Conditions No.6 (IC6):
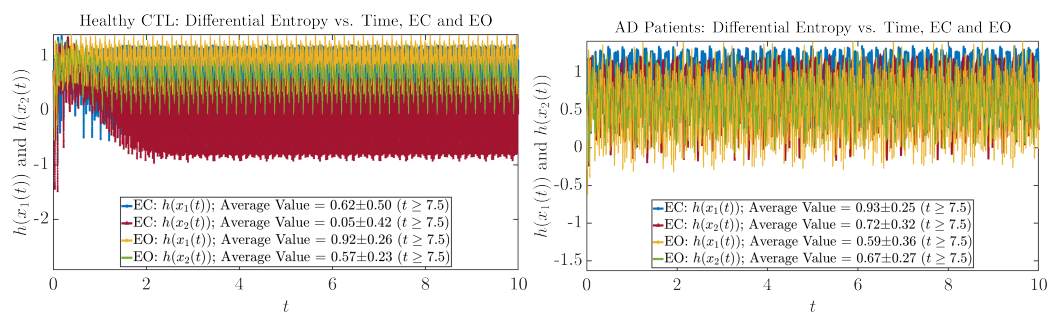
(**a**) IC6: health CTL subjects.

(**b**) IC6: AD patients.

Figure A50. Initial Conditions No.6 (IC6): Shannon differential entropy along time of CTL and AD subjects.

Appendix B.4.2. Empirical Probability Distribution: Shannon Differential Entropy (for $t \geq 7.5$)

(**a**) Initial Conditions No.1 (IC1)

(**b**) Initial Conditions No.2 (IC2)

(**c**) Initial Conditions No.3 (IC3)

(**d**) Initial Conditions No.4 (IC4)

(**e**) Initial Conditions No.5 (IC5)

(**f**) Initial Conditions No.6 (IC6)
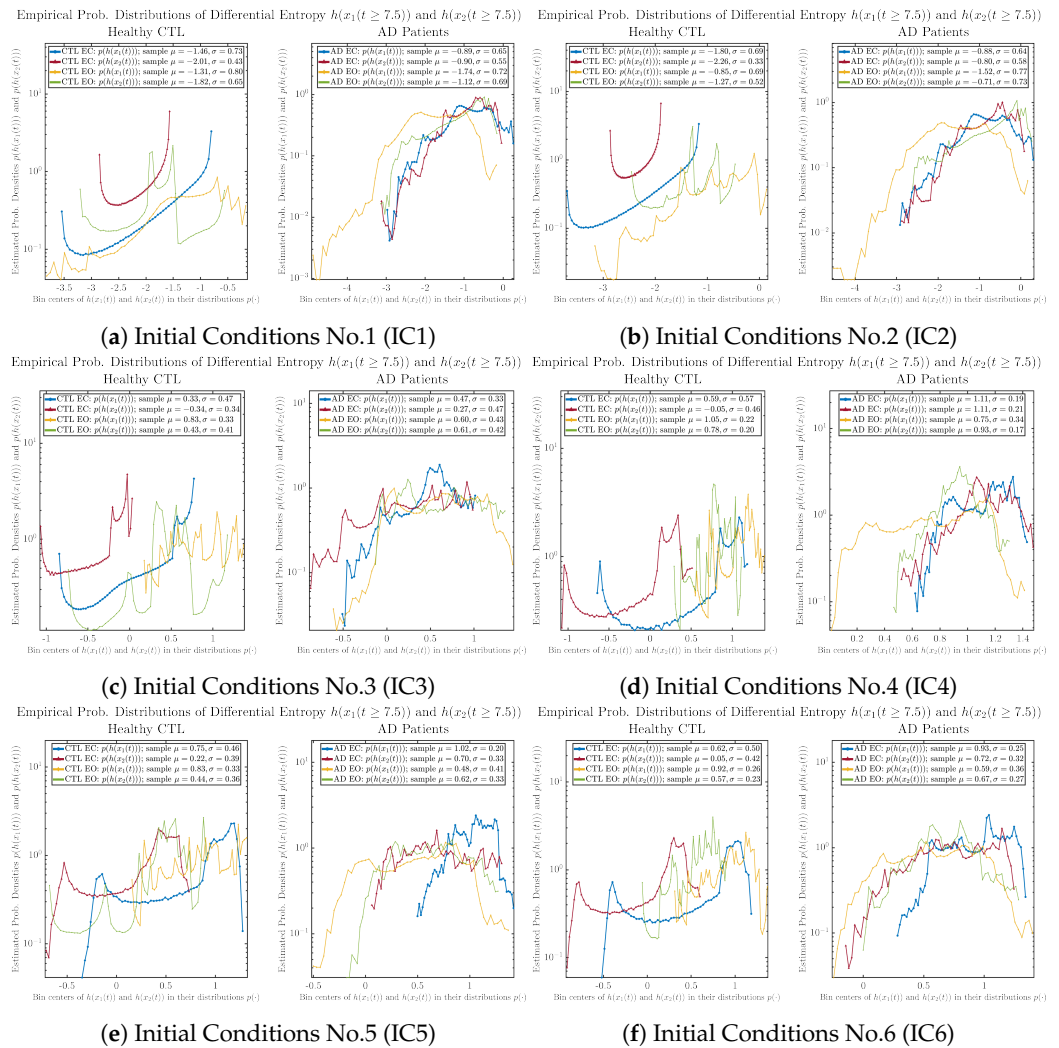
Figure A51. Empirical probability distributions of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$).

Appendix B.4.3. Phase Portraits: Shannon Differential Entropy (for $t \geq 7.5$)

(**a**) Initial Conditions No.1 (IC1)

(**b**) Initial Conditions No.2 (IC2)

(**c**) Initial Conditions No.3 (IC3)

(**d**) Initial Conditions No.4 (IC4)
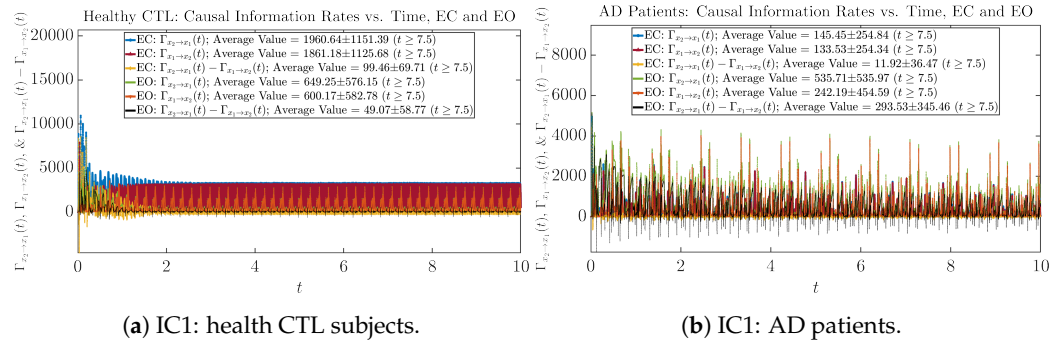
(**e**) Initial Conditions No.5 (IC5)

(**f**) Initial Conditions No.6 (IC6)
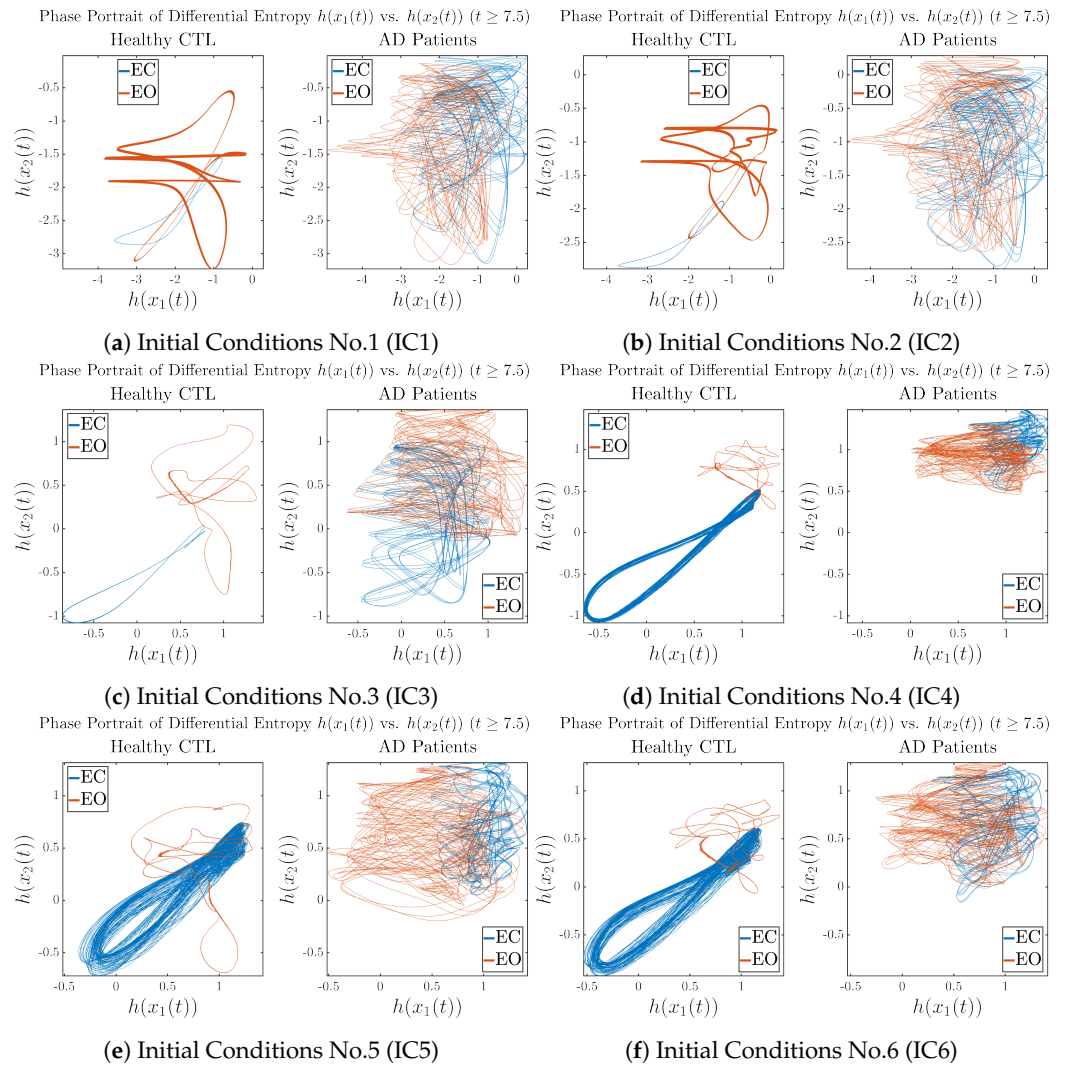
Figure A52. Phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$).

Initial Conditions No.1 (IC1):

(**a**) IC1: health CTL subjects.
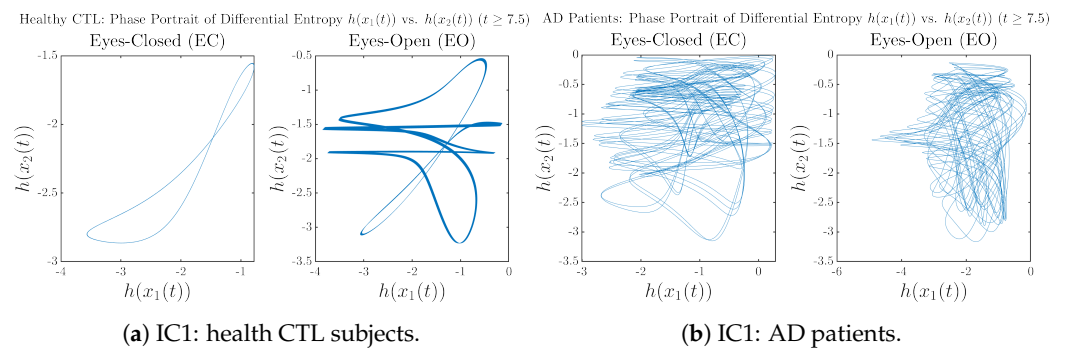
(**b**) IC1: AD patients.

Figure A53. Initial Conditions No.1 (IC1): Phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.2 (IC2):
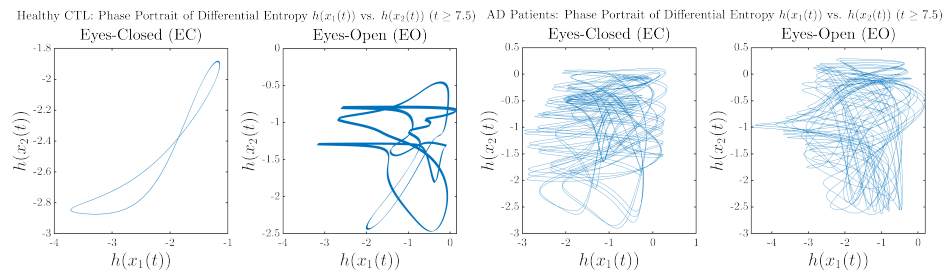
(**a**) IC2: health CTL subjects.　　　　　(**b**) IC2: AD patients.

Figure A54. Initial Conditions No.2 (IC2): Phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.3 (IC3):
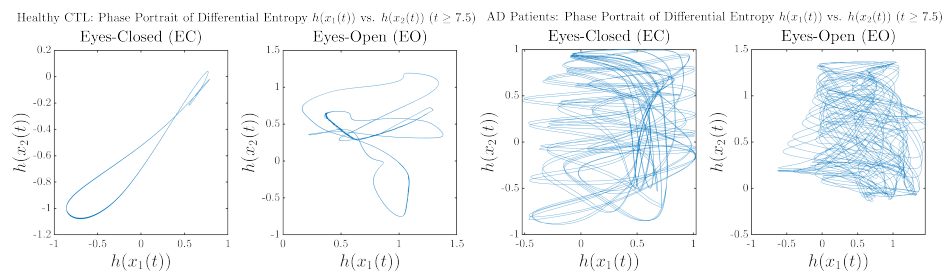
(**a**) IC3: health CTL subjects.　　　　　(**b**) IC3: AD patients.

Figure A55. Initial Conditions No.3 (IC3): Phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.4 (IC4):
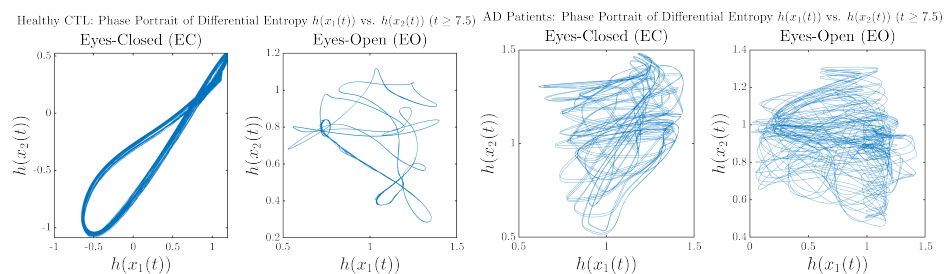
(**a**) IC4: health CTL subjects.　　　　　(**b**) IC4: AD patients.

Figure A56. Initial Conditions No.4 (IC4): Phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.5 (IC5):

(**a**) IC5: health CTL subjects.　　　　　(**b**) IC5: AD patients.

Figure A57. Initial Conditions No.5 (IC5): Phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$) of CTL and AD subjects.

Initial Conditions No.6 (IC6):

(**a**) IC6: health CTL subjects.　　(**b**) IC6: AD patients.

Figure A58. Initial Conditions No.6 (IC6): Phase portraits of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$) of CTL and AD subjects.

Appendix B.4.4. Power Spectra: Shannon Differential Entropy (for $t \geq 7.5$)
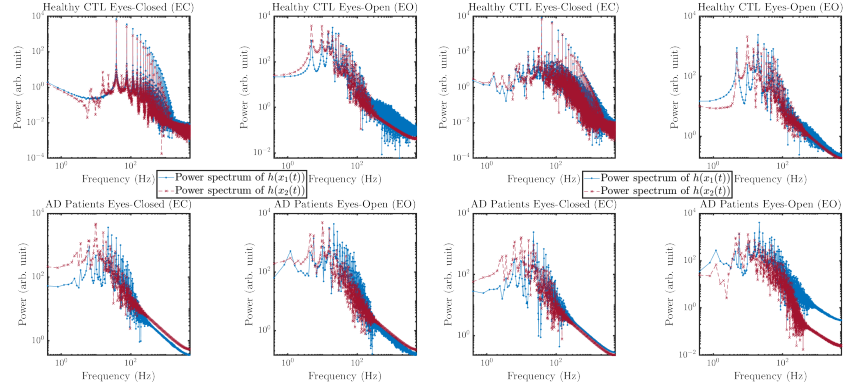
(**a**) Initial Conditions No.1 (IC1)　　(**b**) Initial Conditions No.2 (IC2)

(**c**) Initial Conditions No.3 (IC3)　　(**d**) Initial Conditions No.4 (IC4)
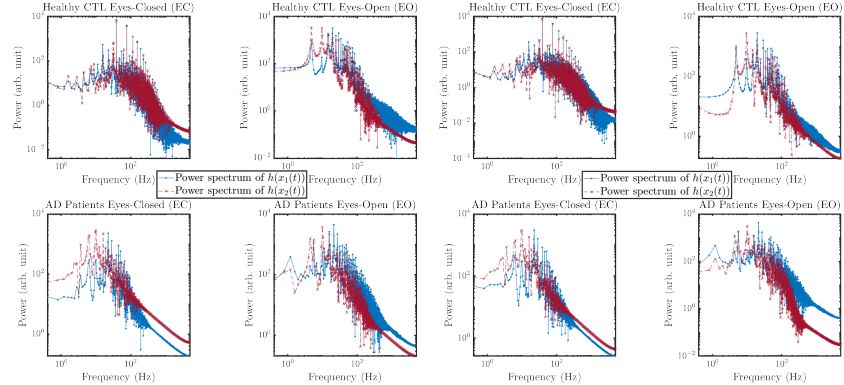
(**e**) Initial Conditions No.5 (IC5)　　(**f**) Initial Conditions No.6 (IC6)

Figure A59. Power spectra of Shannon differential entropy $h(x_1(t))$ and $h(x_2(t))$ ($t \geq 7.5$).

Appendix B.5. Causal Information Rates $\Gamma_{x_2 \to x_1}(t)$, $\Gamma_{x_1 \to x_2}(t)$, *and Net Causal Information Rates* $\Gamma_{x_2 \to x_1}(t) - \Gamma_{x_1 \to x_2}(t)$

Appendix B.5.1. Time Evolution: Causal Information Rates

Initial Conditions No.1 (IC1):

(**a**) IC1: health CTL subjects.

(**b**) IC1: AD patients.

Figure A60. Initial Conditions No.1 (IC1): Causal information rates along time of CTL and AD subjects.
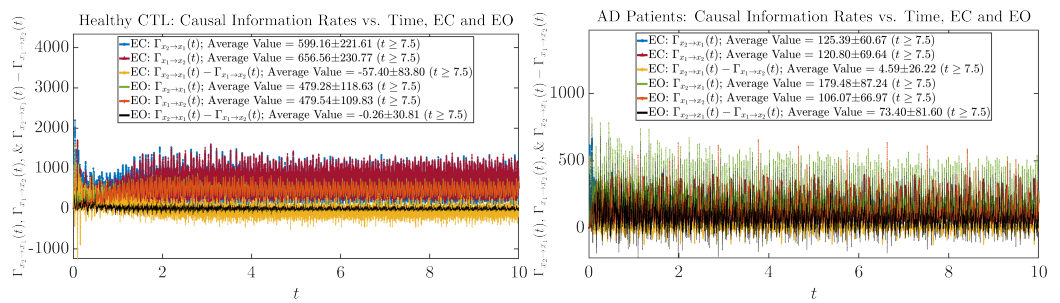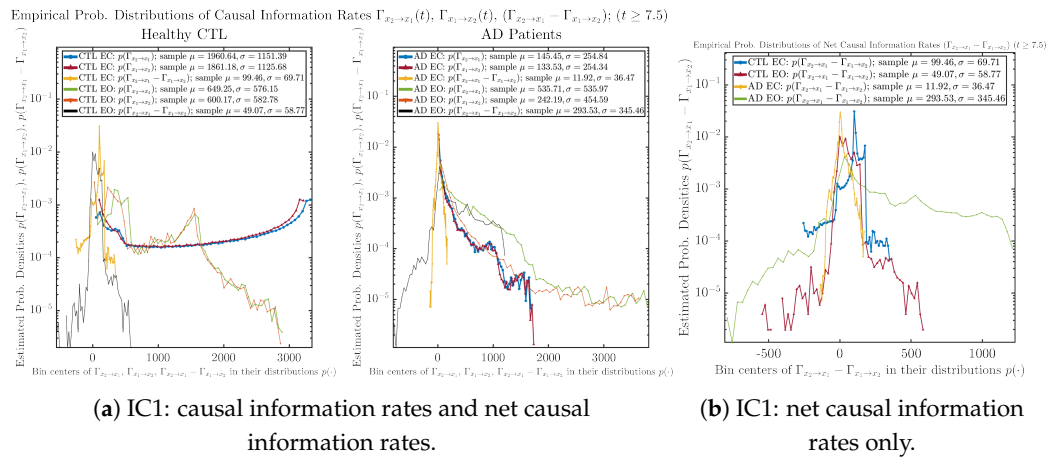
Initial Conditions No.2 (IC2):
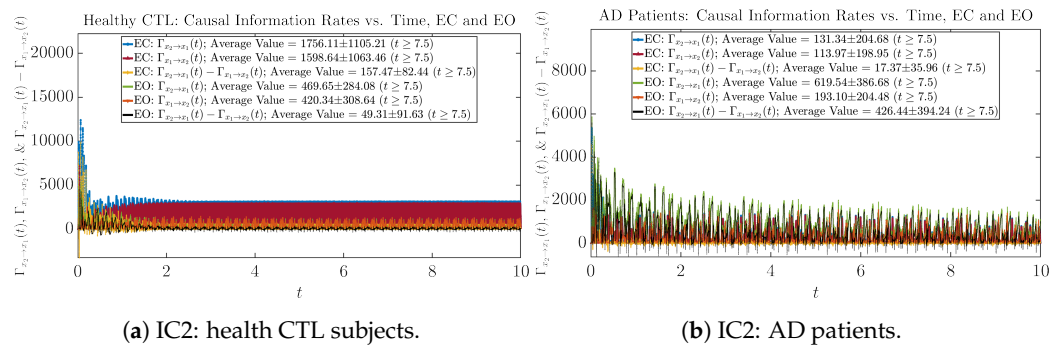
(**a**) IC2: health CTL subjects.

(**b**) IC2: AD patients.

Figure A61. Initial Conditions No.2 (IC2): Causal information rates along time of CTL and AD subjects.

Initial Conditions No.3 (IC3):
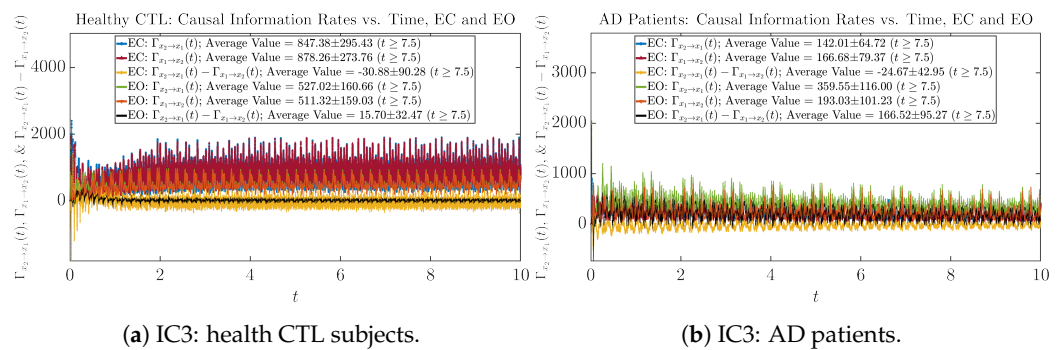
(**a**) IC3: health CTL subjects.

(**b**) IC3: AD patients.

Figure A62. Initial Conditions No.3 (IC3): Causal information rates along time of CTL and AD subjects.

Initial Conditions No.4 (IC4):
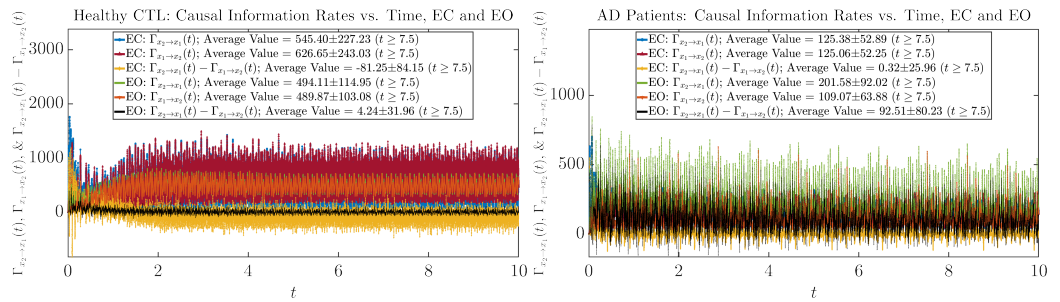
(**a**) IC4: health CTL subjects.　　　　(**b**) IC4: AD patients.

Figure A63. Initial Conditions No.4 (IC4): Causal information rates along time of CTL and AD subjects.

Initial Conditions No.5 (IC5):
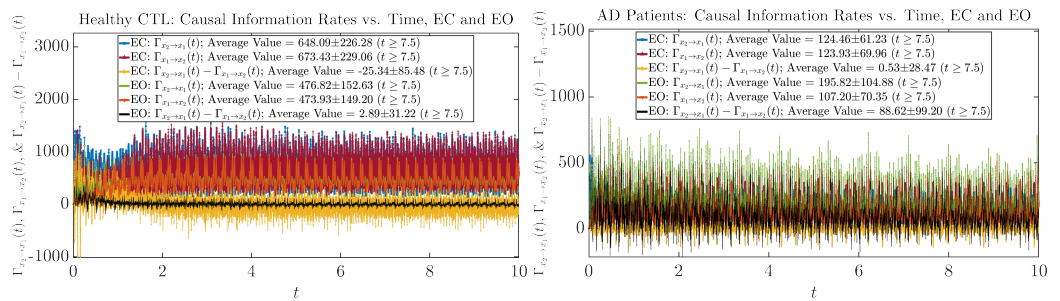
(**a**) IC5: health CTL subjects.　　　　(**b**) IC5: AD patients.

Figure A64. Initial Conditions No.5 (IC5): Causal information rates along time of CTL and AD subjects.

Initial Conditions No.6 (IC6):

(**a**) IC6: health CTL subjects.　　　　(**b**) IC6: AD patients.

Figure A65. Initial Conditions No.6 (IC6): Causal information rates along time of CTL and AD subjects.

Appendix B.5.2. Empirical Probability Distribution: Causal Information Rates (for $t \geq 7.5$)

Initial Conditions No.1 (IC1):

(**a**) IC1: causal information rates and net causal information rates.

(**b**) IC1: net causal information rates only.

Figure A66. Initial Conditions No.1 (IC1): Empirical probability distributions of causal information rates and net causal information rates ($t \geq 7.5$).
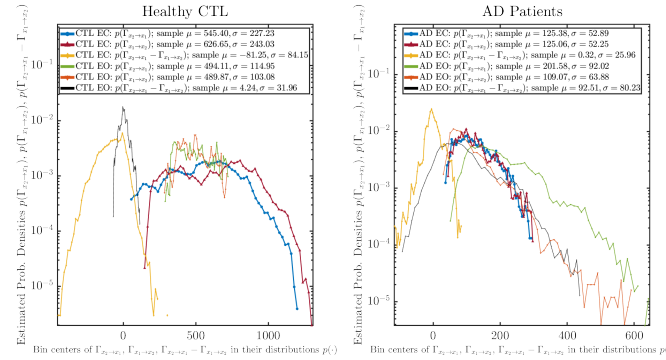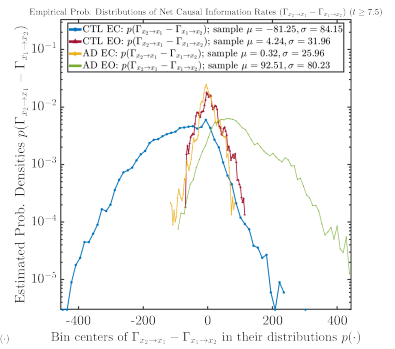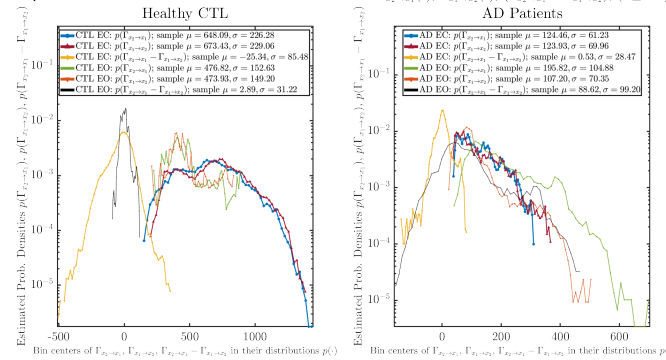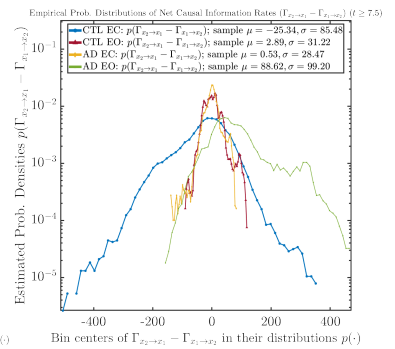
Initial Conditions No.2 (IC2):

(**a**) IC2: causal information rates and net causal information rates.

(**b**) IC2: net causal information rates only.
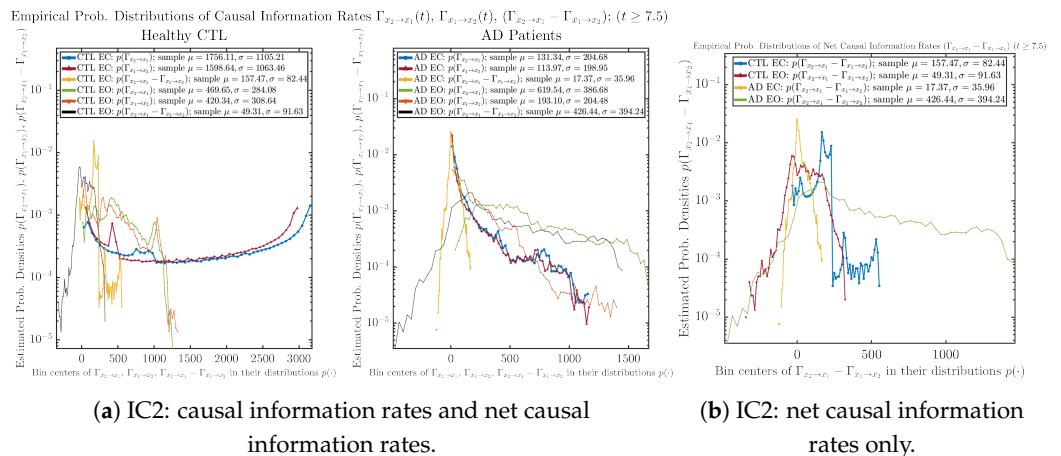
Figure A67. Initial Conditions No.2 (IC2): Empirical probability distributions of causal information rates and net causal information rates ($t \geq 7.5$).

Initial Conditions No.3 (IC3):

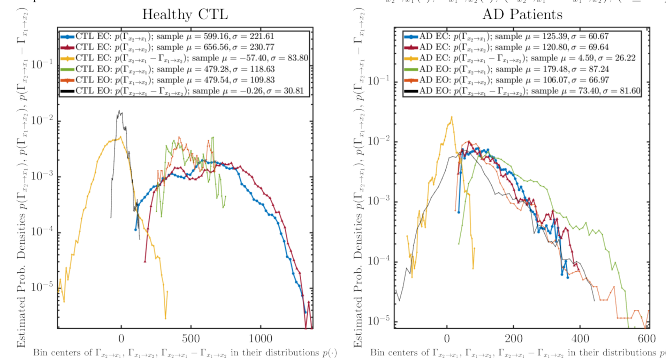(**a**) IC3: causal information rates and net causal information rates.

(**b**) IC3: net causal information rates only.
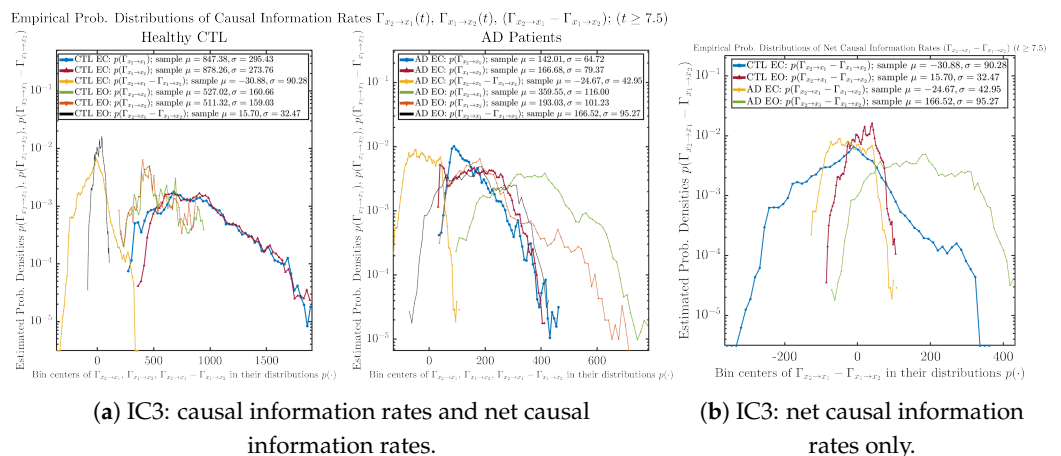
Figure A68. Initial Conditions No.3 (IC3): Empirical probability distributions of causal information rates and net causal information rates ($t \geq 7.5$).

Initial Conditions No.4 (IC4):

(**a**) IC4: causal information rates and net causal information rates.
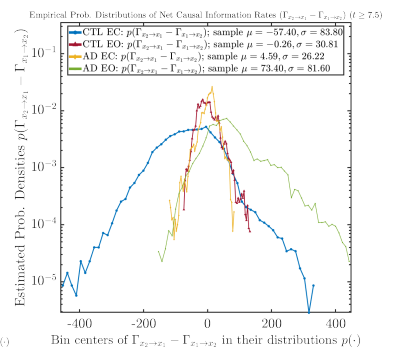
(**b**) IC4: net causal information rates only.

Figure A69. Initial Conditions No.4 (IC4): Empirical probability distributions of causal information rates and net causal information rates ($t \geq 7.5$).

Initial Conditions No.5 (IC5):

(**a**) IC5: causal information rates and net causal information rates.

(**b**) IC5: net causal information rates only.

Figure A70. Initial Conditions No.5 (IC5): Empirical probability distributions of causal information rates and net causal information rates ($t \geq 7.5$).

Initial Conditions No.6 (IC6):

(**a**) IC6: causal information rates and net causal information rates.

(**b**) IC6: net causal information rates only.

Figure A71. Initial Conditions No.6 (IC6): Empirical probability distributions of causal information rates and net causal information rates ($t \geq 7.5$).

Appendix B.6. Causality Based on Transfer Entropy (TE)

Appendix B.6.1. Time Evolution: Transfer Entropy (TE)
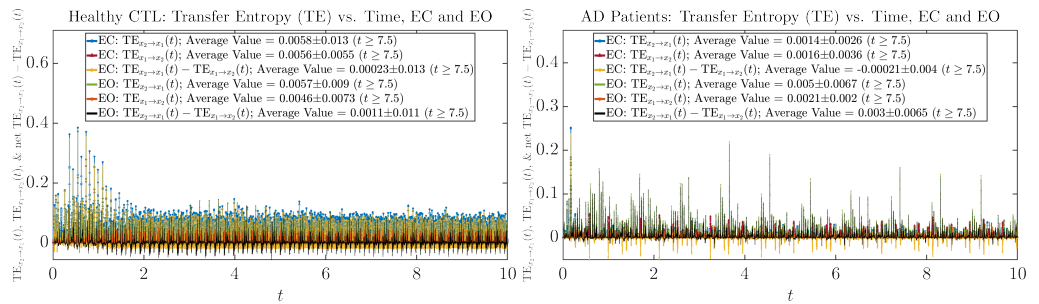
Initial Conditions No.1 (IC1):
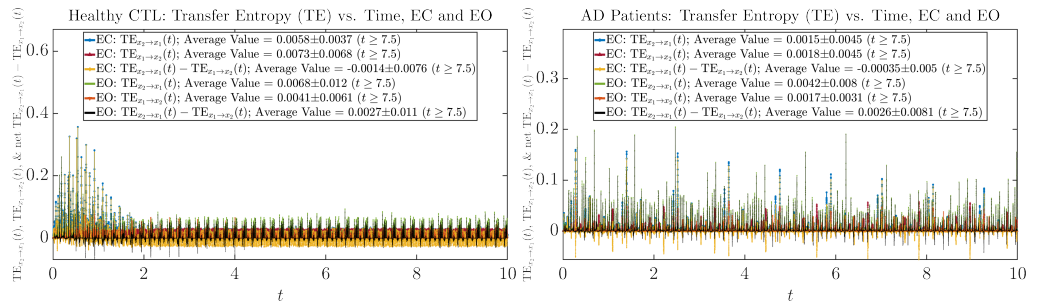
(**a**) IC1: health CTL subjects.   (**b**) IC1: AD patients.

Figure A72. Initial Conditions No.1 (IC1): Transfer Entropy (TE) along time of CTL and AD subjects.

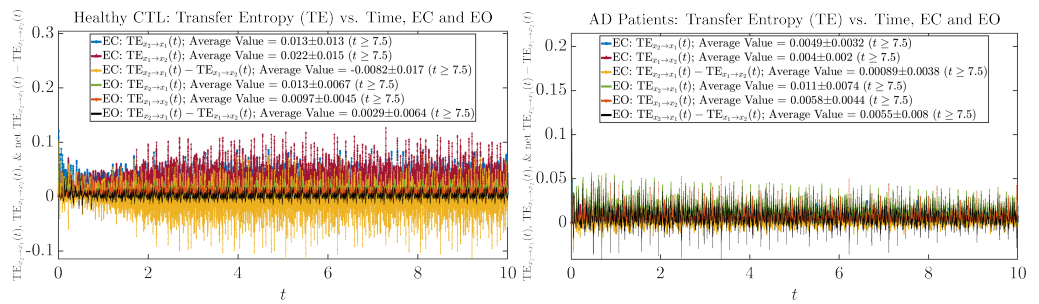Initial Conditions No.2 (IC2):

(**a**) IC2: health CTL subjects.   (**b**) IC2: AD patients.

Figure A73. Initial Conditions No.2 (IC2): Transfer Entropy (TE) along time of CTL and AD subjects.

Initial Conditions No.3 (IC3):

(**a**) IC3: health CTL subjects.   (**b**) IC3: AD patients.

Figure A74. Initial Conditions No.3 (IC3): Transfer Entropy (TE) along time of CTL and AD subjects.
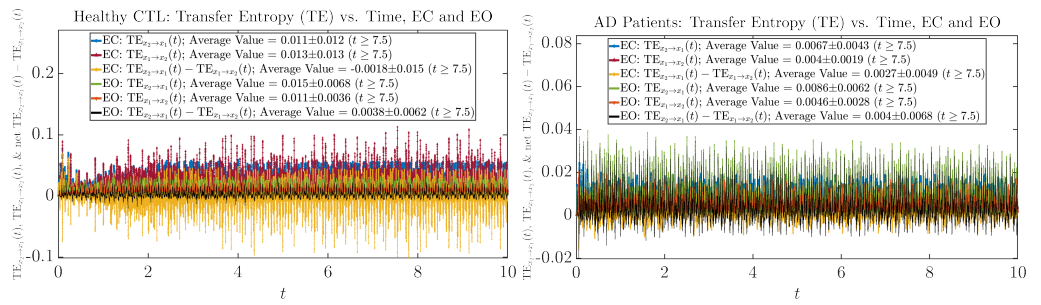
Initial Conditions No.4 (IC4):

(**a**) IC4: health CTL subjects.

(**b**) IC4: AD patients.

Figure A75. Initial Conditions No.4 (IC4): Transfer Entropy (TE) along time of CTL and AD subjects.

Initial Conditions No.5 (IC5):

(**a**) IC5: health CTL subjects.

(**b**) IC5: AD patients.

Figure A76. Initial Conditions No.5 (IC5): Transfer Entropy (TE) along time of CTL and AD subjects.

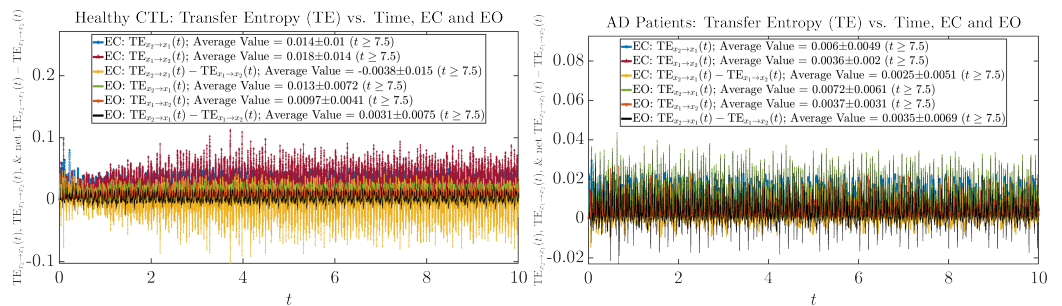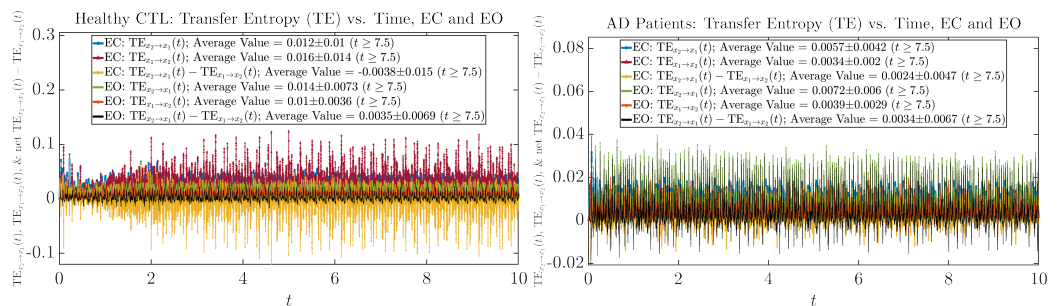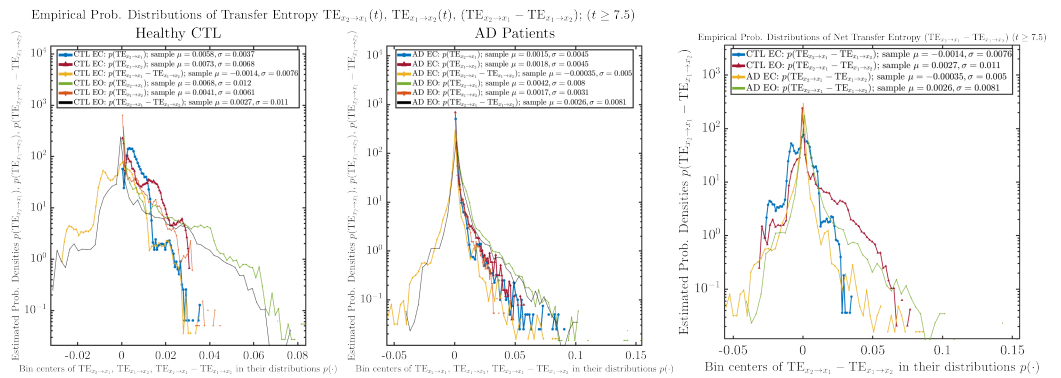Initial Conditions No.6 (IC6):

(**a**) IC6: health CTL subjects.

(**b**) IC6: AD patients.

Figure A77. Initial Conditions No.6 (IC6): Transfer Entropy (TE) along time of CTL and AD subjects.

Appendix B.6.2. Empirical Probability Distribution: Transfer Entropy (TE) (for $t \geq 7.5$)
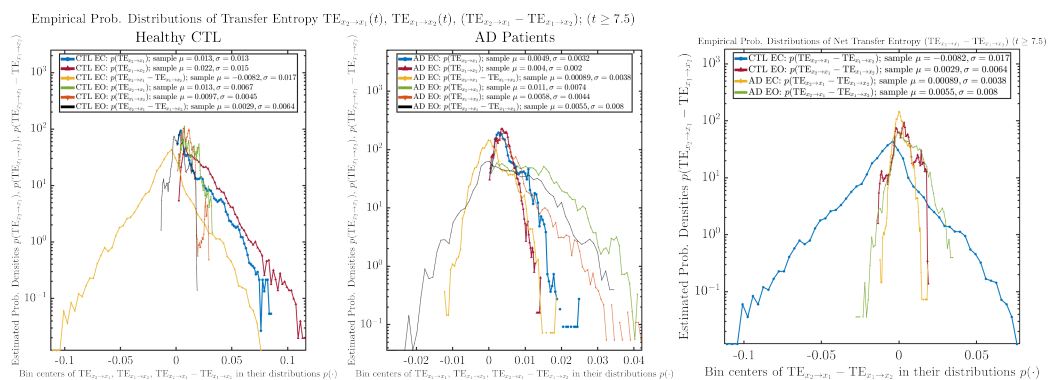
Initial Conditions No.1 (IC1):

(**a**) IC1: Transfer Entropy (TE) and net Transfer Entropy.   (**b**) IC1: Net Transfer Entropy only.

Figure A78. Initial Conditions No.1 (IC1): Empirical probability distributions of Transfer Entropy (TE) and net Transfer Entropy ($t \geq 7.5$).

Initial Conditions No.2 (IC2):
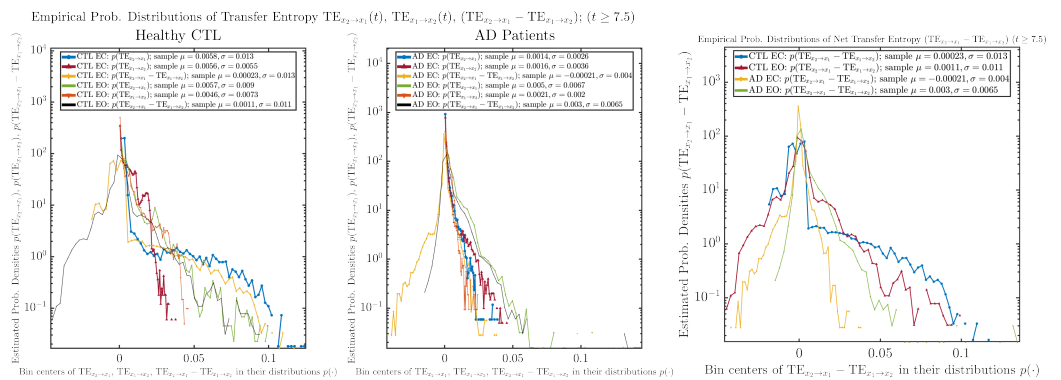
(**a**) IC2: Transfer Entropy (TE) and net Transfer Entropy.   (**b**) IC2: Net Transfer Entropy only.

Figure A79. Initial Conditions No.2 (IC2): Empirical probability distributions of Transfer Entropy (TE) and net Transfer Entropy ($t \geq 7.5$).

Initial Conditions No.3 (IC3):

(**a**) IC3: Transfer Entropy (TE) and net Transfer Entropy.   (**b**) IC3: Net Transfer Entropy only.

Figure A80. Initial Conditions No.3 (IC3): Empirical probability distributions of Transfer Entropy (TE) and net Transfer Entropy ($t \geq 7.5$).

Initial Conditions No.4 (IC4):

(**a**) IC4: Transfer Entropy (TE) and net Transfer Entropy.　(**b**) IC4: Net Transfer Entropy only.

Figure A81. Initial Conditions No.4 (IC4): Empirical probability distributions of Transfer Entropy (TE) and net Transfer Entropy ($t \geq 7.5$).
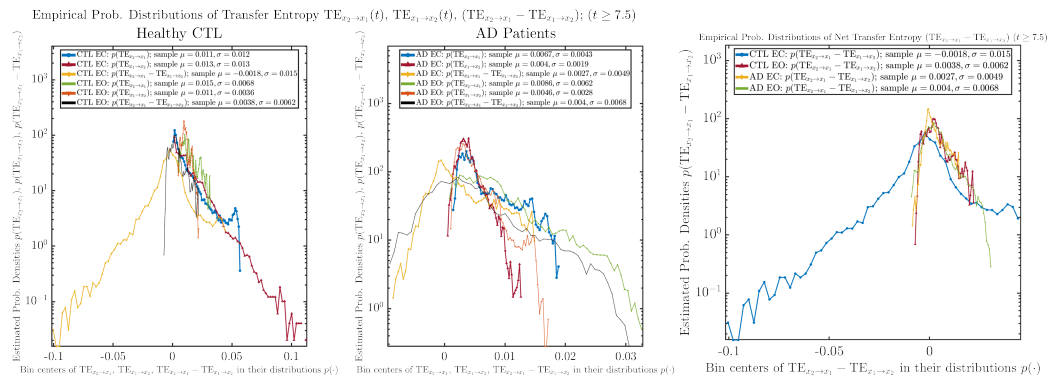
Initial Conditions No.5 (IC5):

(**a**) IC5: Transfer Entropy (TE) and net Transfer Entropy.　(**b**) IC5: Net Transfer Entropy only.

Figure A82. Initial Conditions No.5 (IC5): Empirical probability distributions of Transfer Entropy (TE) and net Transfer Entropy ($t \geq 7.5$).

Initial Conditions No.6 (IC6):

(**a**) IC6: Transfer Entropy (TE) and net Transfer Entropy.　(**b**) IC6: Net Transfer Entropy only.

Figure A83. Initial Conditions No.6 (IC6): Empirical probability distributions of Transfer Entropy (TE) and net Transfer Entropy ($t \geq 7.5$).
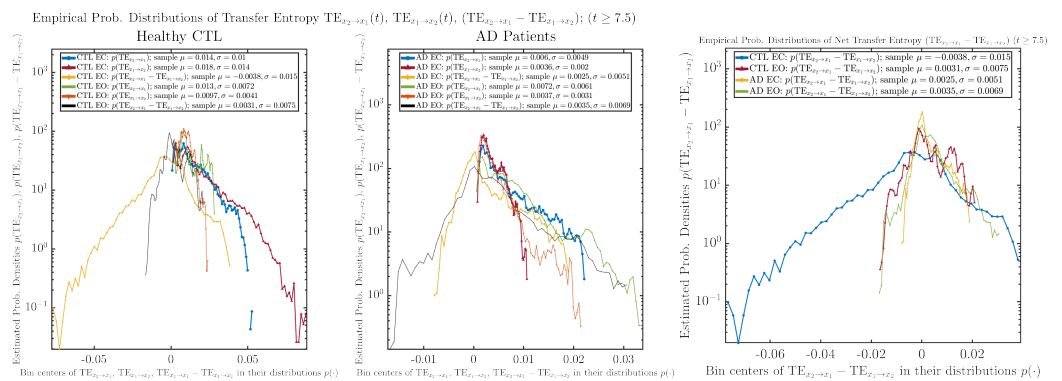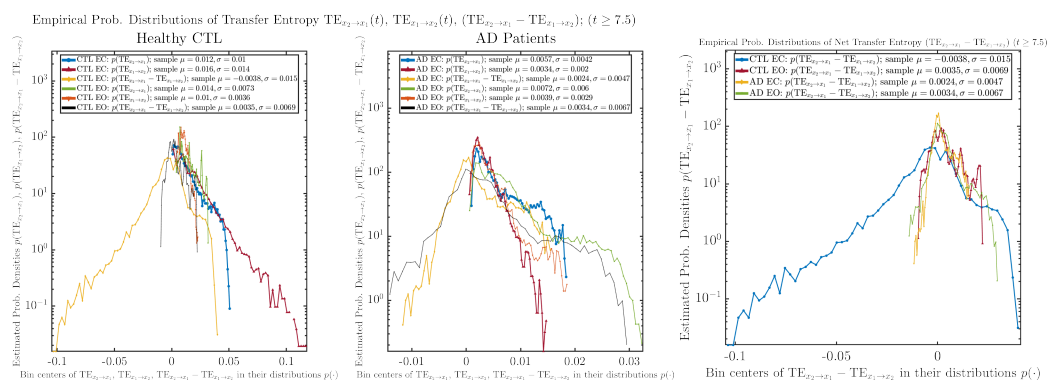
References

1. Ghorbanian, P.; Ramakrishnan, S.; Ashrafiuon, H. Stochastic Non-Linear Oscillator Models of EEG: The Alzheimer's Disease Case. *Front. Comput. Neurosci.* **2015**, *9*, 48. [CrossRef]

2. Szuflitowska, B.; Orlowski, P. Statistical and Physiologically Analysis of Using a Duffing-van Der Pol Oscillator to Modeled Ictal Signals. In Proceedings of the 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV), Shenzhen, China, 13–15 December 2020; pp. 1137–1142. [CrossRef]

3. Nguyen, P.T.M.; Hayashi, Y.; Baptista, M.D.S.; Kondo, T. Collective Almost Synchronization-Based Model to Extract and Predict Features of EEG Signals. *Sci. Rep.* **2020**, *10*, 16342. [CrossRef] [PubMed]

4. Guguloth, S.; Agarwal, V.; Parthasarathy, H.; Upreti, V. Synthesis of EEG Signals Modeled Using Non-Linear Oscillator Based on Speech Data with EKF. *Biomed. Signal Process. Control* **2022**, *77*, 103818. [CrossRef]

5. Szuflitowska, B.; Orlowski, P. Analysis of Parameters Distribution of EEG Signals for Five Epileptic Seizure Phases Modeled by Duffing Van Der Pol Oscillator. In Proceedings of the Computational Science—ICCS 2022, London, UK, 21–23 June 2022; Groen, D., De Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A., Eds.; Springer International Publishing: Cham, Switzerland, 2022; Volume 13352, pp. 188–201. [CrossRef]

6. Barry, R.J.; De Blasio, F.M. EEG Differences between Eyes-Closed and Eyes-Open Resting Remain in Healthy Ageing. *Biol. Psychol.* **2017**, *129*, 293–304. [CrossRef] [PubMed]

7. Jennings, J.L.; Peraza, L.R.; Baker, M.; Alter, K.; Taylor, J.P.; Bauer, R. Investigating the Power of Eyes Open Resting State EEG for Assisting in Dementia Diagnosis. *Alzheimer's Res. Ther.* **2022**, *14*, 109. [CrossRef] [PubMed]

8. Restrepo, J.F.; Mateos, D.M.; López, J.M.D. A Transfer Entropy-Based Methodology to Analyze Information Flow under Eyes-Open and Eyes-Closed Conditions with a Clinical Perspective. *Biomed. Signal Process. Control* **2023**, *86*, 105181. [CrossRef]

9. Klepl, D.; He, F.; Wu, M.; Marco, M.D.; Blackburn, D.J.; Sarrigiannis, P.G. Characterising Alzheimer's Disease with EEG-Based Energy Landscape Analysis. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 992–1000. [CrossRef] [PubMed]

10. Gunawardena, R.; Sarrigiannis, P.G.; Blackburn, D.J.; He, F. Kernel-Based Nonlinear Manifold Learning for EEG-based Functional Connectivity Analysis and Channel Selection with Application to Alzheimer's Disease. *Neuroscience* **2023**, *523*, 140–156. [CrossRef] [PubMed]

11. Barry, R.J.; Clarke, A.R.; Johnstone, S.J.; Magee, C.A.; Rushby, J.A. EEG Differences between Eyes-Closed and Eyes-Open Resting Conditions. *Clin. Neurophysiol.* **2007**, *118*, 2765–2773. [CrossRef]

12. Barry, R.J.; Clarke, A.R.; Johnstone, S.J.; Brown, C.R. EEG Differences in Children between Eyes-Closed and Eyes-Open Resting Conditions. *Clin. Neurophysiol.* **2009**, *120*, 1806–1811. [CrossRef]

13. Matsutomo, N.; Fukami, M.; Kobayashi, K.; Endo, Y.; Kuhara, S.; Yamamoto, T. Effects of Eyes-Closed Resting and Eyes-Open Conditions on Cerebral Blood Flow Measurement Using Arterial Spin Labeling Magnetic Resonance Imaging. *Neurol. Clin. Neurosci.* **2023**, *11*, 10–16. [CrossRef]

14. Agcaoglu, O.; Wilson, T.W.; Wang, Y.P.; Stephen, J.; Calhoun, V.D. Resting State Connectivity Differences in Eyes Open versus Eyes Closed Conditions. *Hum. Brain Mapp.* **2019**, *40*, 2488–2498. [CrossRef] [PubMed]

15. Han, J.; Zhou, L.; Wu, H.; Huang, Y.; Qiu, M.; Huang, L.; Lee, C.; Lane, T.J.; Qin, P. Eyes-Open and Eyes-Closed Resting State Network Connectivity Differences. *Brain Sci.* **2023**, *13*, 122. [CrossRef] [PubMed]

16. Miraglia, F.; Vecchio, F.; Bramanti, P.; Rossini, P.M. EEG Characteristics in "Eyes-Open" versus "Eyes-Closed" Conditions: Small-world Network Architecture in Healthy Aging and Age-Related Brain Degeneration. *Clin. Neurophysiol.* **2016**, *127*, 1261–1268. [CrossRef] [PubMed]

17. Wei, J.; Chen, T.; Li, C.; Liu, G.; Qiu, J.; Wei, D. Eyes-Open and Eyes-Closed Resting States with Opposite Brain Activity in Sensorimotor and Occipital Regions: Multidimensional Evidences from Machine Learning Perspective. *Front. Hum. Neurosci.* **2018**, *12*, 422. [CrossRef] [PubMed]

18. Springer, B.A.; Marin, R.; Cyhan, T.; Roberts, H.; Gill, N.W. Normative Values for the Unipedal Stance Test with Eyes Open and Closed. *J. Geriatr. Phys. Ther.* **2007**, *30*, 8. [CrossRef] [PubMed]

19. Marx, E.; Deutschländer, A.; Stephan, T.; Dieterich, M.; Wiesmann, M.; Brandt, T. Eyes Open and Eyes Closed as Rest Conditions: Impact on Brain Activation Patterns. *NeuroImage* **2004**, *21*, 1818–1824. [CrossRef] [PubMed]

20. Zhang, D.; Liang, B.; Wu, X.; Wang, Z.; Xu, P.; Chang, S.; Liu, B.; Liu, M.; Huang, R. Directionality of Large-Scale Resting-State Brain Networks during Eyes Open and Eyes Closed Conditions. *Front. Hum. Neurosci.* **2015**, *9*, 81. [CrossRef]

21. Xu, P.; Huang, R.; Wang, J.; Van Dam, N.T.; Xie, T.; Dong, Z.; Chen, C.; Gu, R.; Zang, Y.F.; He, Y.; et al. Different Topological Organization of Human Brain Functional Networks with Eyes Open versus Eyes Closed. *NeuroImage* **2014**, *90*, 246–255. [CrossRef]

22. Jeong, J. EEG Dynamics in Patients with Alzheimer's Disease. *Clin. Neurophysiol.* **2004**, *115*, 1490–1505. [CrossRef]

23. Pritchard, W.S.; Duke, D.W.; Coburn, K.L. Altered EEG Dynamical Responsivity Associated with Normal Aging and Probable Alzheimer's Disease. *Dementia* **1991**, *2*, 102–105. [CrossRef]

24. Thiruthummal, A.A.; Kim, E.j. Monte Carlo Simulation of Stochastic Differential Equation to Study Information Geometry. *Entropy* **2022**, *24*, 1113. [CrossRef] [PubMed]

25. Kim, E.j.; Guel-Cortez, A.J. Causal Information Rate. *Entropy* **2021**, *23*, 1087. [CrossRef] [PubMed]

26. Choong, H.J.; Kim, E.j.; He, F. Causality Analysis with Information Geometry: A Comparison. *Entropy* **2023**, *25*, 806. [CrossRef] [PubMed]

27. Baravalle, R.; Rosso, O.A.; Montani, F. Causal Shannon–Fisher Characterization of Motor/Imagery Movements in EEG. *Entropy* **2018**, *20*, 660. [CrossRef] [PubMed]

28. Montani, F.; Baravalle, R.; Montangie, L.; Rosso, O.A. Causal Information Quantification of Prominent Dynamical Features of Biological Neurons. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *373*, 20150109. [CrossRef] [PubMed]

29. Higham, D.J. An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations. *SIAM Rev.* **2001**, *43*, 525–546. [CrossRef]

30. Guel-Cortez, A.-J.; Kim, E.-j. Information Geometric Theory in the Prediction of Abrupt Changes in System Dynamics. *Entropy* **2021**, *23*, 694. [CrossRef] [PubMed]

31. Amari, S.i.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Soc.: Providence, RI, USA, 2000; Volume 191.

32. Frieden, B.R. *Science from Fisher Information*; Cambridge University Press: Cambridge, UK, 2004; Volume 974.

33. Facchi, P.; Kulkarni, R.; Man'ko, V.I.; Marmo, G.; Sudarshan, E.C.G.; Ventriglia, F. Classical and Quantum Fisher Information in the Geometrical Formulation of Quantum Mechanics. *Phys. Lett. A* **2010**, *374*, 4801–4803. [CrossRef]

34. Itoh, M.; Shishido, Y. Fisher Information Metric and Poisson Kernels. *Differ. Geom. Its Appl.* **2008**, *26*, 347–356. [CrossRef]

35. Sahann, R.; Möller, T.; Schmidt, J. Histogram Binning Revisited with a Focus on Human Perception. *arXiv* **2021**, arXiv:cs/2109.06612.

36. Terrell, G.R.; Scott, D.W. Oversmoothed Nonparametric Density Estimates. *J. Am. Stat. Assoc.* **1985**, *80*, 209–214. [CrossRef]

37. Budišić, M.; Mohr, R.M.; Mezić, I. Applied Koopmanism. *Chaos Interdiscip. J. Nonlinear Sci.* **2012**, *22*, 047510. [CrossRef] [PubMed]

38. Koopman, B.O. Hamiltonian Systems and Transformation in Hilbert Space. *Proc. Natl. Acad. Sci. USA* **1931**, *17*, 315–318. [CrossRef]