Research

Author for correspondence:
Naoki Masuda
e-mail: naoki.masuda@bristol.ac.uk

THE ROYAL SOCIETY
PUBLISHING

Energy landscape analysis of neuroimaging data

Takahiro Ezaki[1,2], Takamitsu Watanabe[3], Masayuki Ohzeki[4] and Naoki Masuda[5]

[1]National Institute of Informatics, Hitotsubashi, Chiyoda-ku, Tokyo, Japan
[2]Kawarabayashi Large Graph Project, ERATO, JST, c/o Global Research Center for Big Data Mathematics, NII, Chiyoda-ku, Tokyo, Japan
[3]Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London WC1N 3AZ, UK
[4]Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan
[5]Department of Engineering Mathematics, University of Bristol, Bristol BS8 1UB, UK

NM, 0000-0003-1567-801X

Computational neuroscience models have been used for understanding neural dynamics in the brain and how they may be altered when physiological or other conditions change. We review and develop a data-driven approach to neuroimaging data called the energy landscape analysis. The methods are rooted in statistical physics theory, in particular the Ising model, also known as the (pairwise) maximum entropy model and Boltzmann machine. The methods have been applied to fitting electrophysiological data in neuroscience for a decade, but their use in neuroimaging data is still in its infancy. We first review the methods and discuss some algorithms and technical aspects. Then, we apply the methods to functional magnetic resonance imaging data recorded from healthy individuals to inspect the relationship between the accuracy of fitting, the size of the brain system to be analysed and the data length.

This article is part of the themed issue 'Mathematical methods in medicine: neuroscience, cardiology and pathology'.

1. Introduction

Altered cognitive function due to various neuropsychiatric disorders seems to result from aberrant neural dynamics in the affected brain [1–4]. Alterations in brain dynamics may also occur in the absence of disorders, in situations such as typical ageing, traumatic experiences, emotional responses and tasks. Functional magnetic resonance imaging (fMRI) provides information on the neural dynamics in the brain with a reasonable spatial resolution in a non-invasive manner. There are various analysis methods that can be used to extract the dynamics in neuroimaging data including fMRI signals, such as sliding-window functional connectivity analysis, dynamic causal modelling, oscillation analysis and biophysical modelling. In this study, we seek the potential of a different approach: energy landscape analysis.

This method is rooted in statistical physics. The main idea is to map the brain dynamics to the movement of a 'ball' constrained on an energy landscape inferred from neural data. A ball tends to go downhill and remains near the bottom of a basin in a landscape, whereas it sometimes goes uphill due to random fluctuations that cause it to wander around and possibly transit to another basin (figure 1*g*). Using the Ising model (equivalent to the Boltzmann machine and the pairwise maximum entropy model (MEM); see [5–7] for reviews in neuroscience), we can explicitly construct an energy landscape from multivariate time-series data including fMRI signals recorded at a specified set of regions of interest (ROIs). The pairwise MEM, or, equivalently, the Ising model, has been used to emulate fMRI signals [6,8–11]. More recently, we have used the pairwise MEM for fMRI data during rest [12] and sleep [13] and then developed an energy landscape analysis method and applied it to participants during rest [14] and during a bistable visual perception task [15]. In contrast with the aforementioned previous studies [6,8–11], our approach is data driven, with the parameters of the Ising model being inferred from the given data. In this paper, we review the methods and some technical details. In passing, we introduce new techniques (i.e. different inference algorithms and a Dijkstra-like method). We apply the methods to publicly shared resting-state fMRI data recorded from healthy human participants to validate the new approaches and also to examine the relationship between the accuracy of fit, the size of the brain system (i.e. number of ROIs) and the length of the fMRI data.

2. Material and methods

The pipeline of the energy landscape analysis based on the pairwise MEM is illustrated in figure 1.

(a) Pairwise maximum entropy model

First, we specify a brain network of interest (figure 1*a*) and obtain resting-state (or under other conditions, which are ideally stationary) fMRI signals at the ROIs, resulting in a multivariate time series (figure 1*b*). We denote the number of ROIs by N.

Second, we binarize the fMRI signal at each time point (i.e. in each image volume) and each ROI by thresholding the signal. Then, for each ROI i ($i = 1, \ldots, N$), we obtain a sequence of binarized signals representing the brain activity, $\{\sigma_i(1), \ldots, \sigma_i(t_{\max})\}$, where $t_{\max}$ is the length of the data, $\sigma_i(t) = 1$ ($t = 1, \ldots, t_{\max}$) indicates that the ith ROI is active at time t, and $\sigma_i(t) = -1$ indicates that the ROI is inactive (figure 1*c*). The threshold is arbitrary, and we set it to the time average of $\sigma_i(t)$ for each i. The activity pattern of the entire network at time t is given by an N-dimensional vector $\boldsymbol{\sigma} \equiv (\sigma_1, \ldots, \sigma_N) \in \{-1, 1\}^N$, where we have suppressed t. Note that there are 2^N possible activity patterns in total. Binarization is not readily justified given continuously distributed fMRI signals. However, we previously showed that the pairwise MEM with binarized signals predicted anatomical connectivity of the brain better than other functional connectivity methods that are based on non-binarized continuous fMRI signals and that ternary as opposed to binary quantization did not help to improve the results [12].
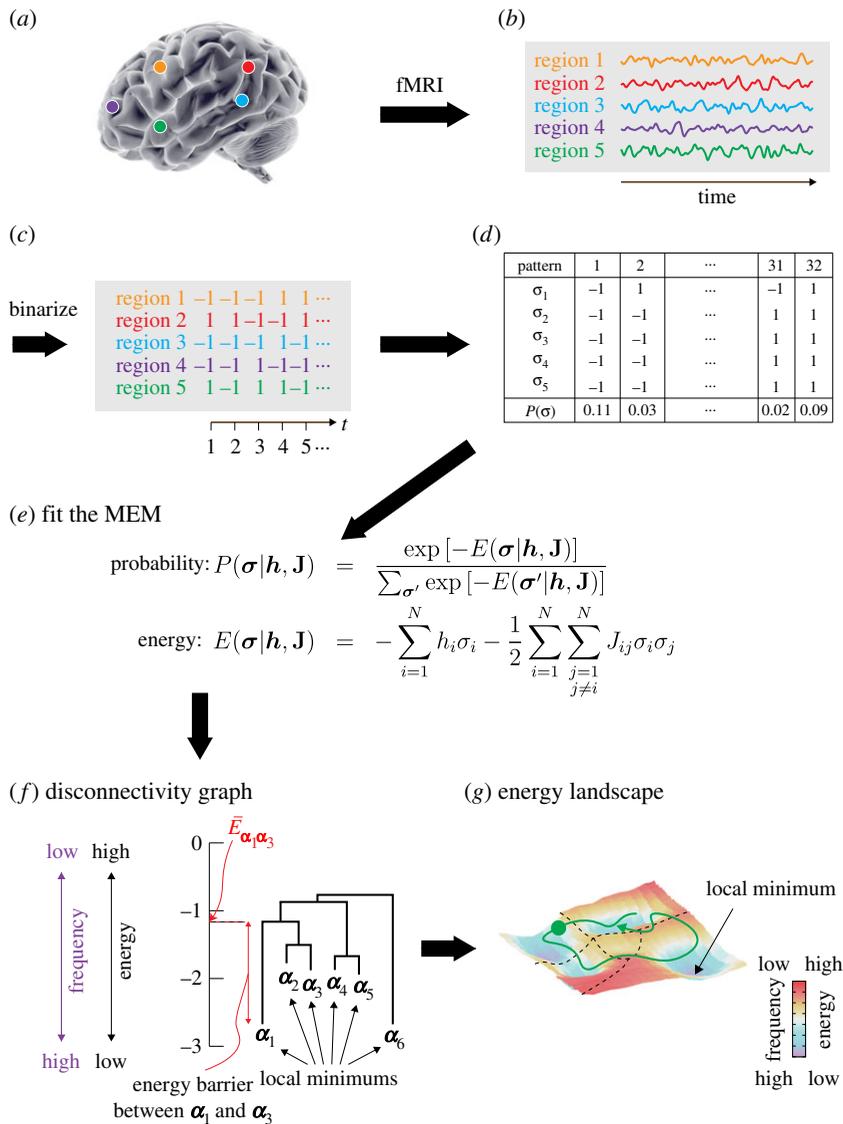
Figure 1. Procedures of the energy landscape analysis for fMRI data. (*a*) ROIs are specified. (*b*) fMRI signals are measured at the ROIs. (*c*) The fMRI signal at each ROI and each time point is binarized into 1 (active) or −1 (inactive). (*d*) The normalized frequency is computed for each binarized activity pattern. (*e*) The pairwise MEM model (i.e. Boltzmann distribution) is fitted to the empirical distribution of the 2^N activity patterns (equation (2.1)). The energy value is also obtained for each activity pattern (equation (2.2)). (*f*) Relationships between activity patterns that are energy local minimums are summarized into a disconnectivity graph. (*g*) Schematic of the energy landscape. Each local minimum corresponds to the bottom of a basin. The borders between attractive basins of different local minimums are shown by the dotted curves. Any activity pattern belongs to the basin of a local minimum. Brain dynamics can be interpreted as the motion of a 'ball' constrained on the energy landscape. (Online version in colour.)

Third, we calculate the relative frequency with which each activity pattern is visited, $P_{\text{empirical}}(\boldsymbol{\sigma})$ (figure 1*d*). To $P_{\text{empirical}}(\boldsymbol{\sigma})$, we fit the Boltzmann distribution given by

$$P(\boldsymbol{\sigma} \mid \boldsymbol{h}, \mathbf{J}) = \frac{\exp\left[-E(\boldsymbol{\sigma} \mid \boldsymbol{h}, \mathbf{J})\right]}{\sum_{\boldsymbol{\sigma}'} \exp\left[-E(\boldsymbol{\sigma}' \mid \boldsymbol{h}, \mathbf{J})\right]}, \tag{2.1}$$

where

$$E(\boldsymbol{\sigma} \,|\, \boldsymbol{h}, \mathbf{J}) = - \sum_{i=1}^{N} h_i \sigma_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} J_{ij} \sigma_i \sigma_j \tag{2.2}$$

is the energy, and $\boldsymbol{h} = \{h_i\}$ and $\mathbf{J} = \{J_{ij}\}$ $(i, j = 1, \ldots, N)$ are the parameters of the model (figure 1e). We assume $J_{ij} = J_{ji}$ and $J_{ii} = 0$ $(i, j = 1, \ldots, N)$. The principle of maximum entropy imposes that we select $\boldsymbol{h}$ and $\mathbf{J}$ such that $\langle \sigma_i \rangle_{\text{empirical}} = \langle \sigma_i \rangle_{\text{model}}$ and $\langle \sigma_i \sigma_j \rangle_{\text{empirical}} = \langle \sigma_i \sigma_j \rangle_{\text{model}}$ $(i, j = 1, \ldots, N)$, where $\langle \cdots \rangle_{\text{empirical}}$ and $\langle \cdots \rangle_{\text{model}}$ represent the mean with respect to the empirical distribution (figure 1d) and the model distribution (equation (2.1)), respectively. By maximizing the entropy of the distribution under these constraints, we obtain the Boltzmann distribution given by equation (2.1). Some algorithms for the fitting will be explained in §2b. Equation (2.1) indicates that an activity pattern with a high energy value is not frequently visited and vice versa. Values of h_i and J_{ij} represent the baseline activity at the ith ROI and the interaction between the ith and jth ROIs, respectively. Equation (2.2) implies that, if h_i is large, the energy is smaller with $\sigma_i = 1$ than with $\sigma_i = -1$, such that the ith ROI tends to be active.

(b) Algorithms to estimate the pairwise maximum entropy model

In this section, we review three algorithms to estimate the parameters of the MEM, i.e. $\boldsymbol{h}$ and $\mathbf{J}$.

(i) Likelihood maximization

In the maximum-likelihood method, we solve

$$(\boldsymbol{h}, \mathbf{J}) = \arg\max_{\boldsymbol{h}, \mathbf{J}} \mathcal{L}(\boldsymbol{h}, \mathbf{J}), \tag{2.3}$$

where $\mathcal{L}(\boldsymbol{h}, \mathbf{J})$ is the likelihood given by

$$\mathcal{L}(\boldsymbol{h}, \mathbf{J}) = \prod_{t=1}^{t_{\max}} P(\boldsymbol{\sigma}(t) \,|\, \boldsymbol{h}, \mathbf{J}). \tag{2.4}$$

We can maximize the likelihood by a gradient ascent scheme

$$h_i^{\text{new}} - h_i^{\text{old}} = \frac{\epsilon}{t_{\max}} \frac{\partial}{\partial h_i} \log \mathcal{L}(\boldsymbol{h}, \mathbf{J}) = \epsilon(\langle \sigma_i \rangle_{\text{empirical}} - \langle \sigma_i \rangle_{\text{model}}) \tag{2.5}$$

and

$$J_{ij}^{\text{new}} - J_{ij}^{\text{old}} = \frac{\epsilon}{t_{\max}} \frac{\partial}{\partial J_{ij}} \log \mathcal{L}(\boldsymbol{h}, \mathbf{J}) = \epsilon(\langle \sigma_i \sigma_j \rangle_{\text{empirical}} - \langle \sigma_i \sigma_j \rangle_{\text{model}}), \tag{2.6}$$

where the superscripts new and old represent the values after and before a single updating step, respectively, and ϵ (>0) is a constant. A slightly different updating scheme called the iterative scaling algorithm [16], where the right-hand side of equations (2.5) and (2.6) is replaced by $\epsilon \, \text{sgn}(\langle \sigma_i \rangle) \log(\langle \sigma_i \rangle_{\text{empirical}}/\langle \sigma_i \rangle_{\text{model}})$ and $\epsilon \, \text{sgn}(\langle \sigma_i \sigma_j \rangle) \log(\langle \sigma_i \sigma_j \rangle_{\text{empirical}}/\langle \sigma_i \sigma_j \rangle_{\text{model}})$, respectively, is also used sometimes [5,12,17,18]. Because equation (2.1) is concave in terms of $\boldsymbol{h}$ and $\mathbf{J}$ (which we can show by verifying that the Hessian of $\log \mathcal{L}$ is a type of sign-flipped covariance matrix, which is negative semi-definite), the gradient ascent scheme yields the maximum-likelihood estimator. Because a single updating step involves all the 2^N activity patterns to calculate $\langle \sigma_i \rangle_{\text{model}}$ and $\langle \sigma_i \sigma_j \rangle_{\text{model}}$, likelihood maximization is computationally costly for large N.

Our Matlab code to calculate the maximum-likelihood estimator for arbitrary multivariate time-series data is available in the electronic supplementary material.

(ii) Pseudo-likelihood maximization

The pseudo-likelihood maximization method approximates the likelihood function as follows:

$$\mathcal{L}(\boldsymbol{h}, \mathbf{J}) \approx \prod_{t=1}^{t_{\max}} \prod_{i=1}^{N} \tilde{P}(\sigma_i \mid \boldsymbol{h}, \mathbf{J}, \boldsymbol{\sigma}_{/i}(t)), \tag{2.7}$$

where $\tilde{P}$ represents the Boltzmann distribution for a single spin, σ_i, given the other $\sigma_j (j \neq i)$ values fixed to $\boldsymbol{\sigma}_{/i}(t) \equiv (\sigma_1(t), \dots, \sigma_{i-1}(t), \sigma_{i+1}(t), \dots, \sigma_N(t))$ [19]. In other words,

$$\tilde{P}(\sigma_i | \boldsymbol{h}, \mathbf{J}, \boldsymbol{\sigma}_{/i}(t)) = \frac{\exp[h_i \sigma_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} J_{ij} \sigma_i \sigma_j(t)]}{\sum_{\sigma_i' = -1, +1} \exp[h_i \sigma_i' + \sum_{\substack{j=1 \\ j \neq i}}^{N} J_{ij} \sigma_i' \sigma_j(t)]}. \tag{2.8}$$

We call the right-hand side of equation (2.7) the pseudo-likelihood. Although this is a mean-field approximation neglecting the influence of σ_i on σ_j $(j \neq i)$, the estimator obtained by the maximization of the pseudo-likelihood approaches the maximum-likelihood estimator as $t_{\max} \to \infty$ [19]. A gradient ascent updating scheme is given by

$$h_i^{\text{new}} - h_i^{\text{old}} = \epsilon(\langle \sigma_i \rangle_{\text{empirical}} - \langle \sigma_i \rangle_{\tilde{P}}) \tag{2.9}$$

and

$$J_{ij}^{\text{new}} - J_{ij}^{\text{old}} = \epsilon(\langle \sigma_i \sigma_j \rangle_{\text{empirical}} - \langle \sigma_i \sigma_j \rangle_{\tilde{P}}), \tag{2.10}$$

where $\langle \sigma_i \rangle_{\tilde{P}}$ and $\langle \sigma_i \sigma_j \rangle_{\tilde{P}}$ are the mean and correlation with respect to distribution $\tilde{P}$ (equation (2.8)) and are given by

$$\langle \sigma_i \rangle_{\tilde{P}} = \frac{1}{t_{\max}} \sum_{t=1}^{t_{\max}} \tanh \left[h_i + \sum_{\substack{j'=1 \\ j' \neq i}}^{N} J_{ij'} \sigma_{j'}(t) \right] \tag{2.11}$$

and

$$\langle \sigma_i \sigma_j \rangle_{\tilde{P}} = \frac{1}{t_{\max}} \sum_{t=1}^{t_{\max}} \sigma_j(t) \tanh \left[h_i + \sum_{\substack{j'=1 \\ j' \neq i}}^{N} J_{ij'} \sigma_{j'}(t) \right], \tag{2.12}$$

respectively. It should be noted that this updating rule circumvents the calculation of $\langle \sigma_i \rangle_{\text{model}}$ and $\langle \sigma_i \sigma_j \rangle_{\text{model}}$, which the gradient ascent method to maximize the original likelihood uses and involves 2^N terms.

(iii) Minimum probability flow

Different from the likelihood and pseudo-likelihood maximization, the minimum probability flow method [20] is not based on the likelihood function. Consider relaxation dynamics of a probability distribution, $P(\boldsymbol{\sigma}; \tau)$, on the 2^N activity patterns whose master equation is given by

$$\frac{\mathrm{d}P(\boldsymbol{\sigma}; \tau)}{\mathrm{d}\tau} = \sum_{\boldsymbol{\sigma}'} [W(\boldsymbol{\sigma} \mid \boldsymbol{\sigma}') P(\boldsymbol{\sigma}'; \tau) - W(\boldsymbol{\sigma}' \mid \boldsymbol{\sigma}) P(\boldsymbol{\sigma}; \tau)], \tag{2.13}$$

where $W(\boldsymbol{\sigma}|\boldsymbol{\sigma}')$ is a transition rate from activity pattern $\boldsymbol{\sigma}'$ to activity pattern $\boldsymbol{\sigma}$. As the initial condition, we impose $P(\boldsymbol{\sigma}; 0) = P_{\text{empirical}}(\boldsymbol{\sigma})$. By choosing

$$W(\boldsymbol{\sigma}|\boldsymbol{\sigma}') = \begin{cases} \exp[-\frac{1}{2}(E(\boldsymbol{\sigma}|\boldsymbol{h}, \mathbf{J}) - E(\boldsymbol{\sigma}'|\boldsymbol{h}, \mathbf{J}))] & (\boldsymbol{\sigma} \text{ and } \boldsymbol{\sigma}' \text{ are neighbouring patterns}), \\ 0 & (\text{otherwise}), \end{cases} \tag{2.14}$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ are neighbours if they are only different at one ROI, we obtain a standard Markov chain Monte Carlo method such that $P(\boldsymbol{\sigma}; \tau)$ converges to the Boltzmann distribution given by equation (2.1).

In the minimum probability flow method, we look for h and J values for which $P(\boldsymbol{\sigma};\tau)$ changes little in the relaxation dynamics at $\tau = 0$ [20]. The idea is that only a small amount of relaxation is necessary if the initial distribution, i.e. $P(\boldsymbol{\sigma};0)$ ($=P_{\mathrm{empirical}}(\boldsymbol{\sigma})$), is sufficiently close to the equilibrium distribution, i.e. the Boltzmann distribution. The Kullback–Leibler (KL) divergence between the empirical distribution, $P(\boldsymbol{\sigma};0)$, and a probability distribution after an infinitely small relaxation time, $P(\boldsymbol{\sigma};\epsilon)$, is approximated as:

$$
D_{\mathrm{KL}}(P(\boldsymbol{\sigma};0) \parallel P(\boldsymbol{\sigma};\epsilon)) \approx D_{\mathrm{KL}}(P(\boldsymbol{\sigma};0) \parallel P(\boldsymbol{\sigma};\tau))|_{\tau=0} + \epsilon \left. \frac{\mathrm{d}D_{\mathrm{KL}}(P(\boldsymbol{\sigma};0) \parallel P(\boldsymbol{\sigma};\tau))}{\mathrm{d}\tau} \right|_{\tau=0}
$$

$$
= \epsilon \frac{\mathrm{d}}{\mathrm{d}\tau} \left[\sum_{\boldsymbol{\sigma} \in \mathcal{D}} P(\boldsymbol{\sigma};0) \log \frac{P(\boldsymbol{\sigma};0)}{P(\boldsymbol{\sigma};\tau)} \right]\Bigg|_{\tau=0} = -\epsilon \sum_{\boldsymbol{\sigma} \in \mathcal{D}} \left. \frac{\mathrm{d}P(\boldsymbol{\sigma};\tau)}{\mathrm{d}\tau} \right|_{\tau=0}
$$

$$
= \frac{\epsilon}{t_{\max}} \sum_{t=1}^{t_{\max}} \sum_{\boldsymbol{\sigma}' \in \Omega \setminus \mathcal{D}} W(\boldsymbol{\sigma}' \mid \boldsymbol{\sigma}(t)), \tag{2.15}
$$

where $D_{\mathrm{KL}}(P(\boldsymbol{\sigma};0) \parallel P(\boldsymbol{\sigma};\epsilon)) = \sum_{\boldsymbol{\sigma} \in \Omega} P(\boldsymbol{\sigma};0) \log[P(\boldsymbol{\sigma};0)/P(\boldsymbol{\sigma};\epsilon)]$ is the KL divergence, which quantifies the discrepancy between two distributions, Ω is the set of all the 2^N activity patterns and $\mathcal{D} = \{\boldsymbol{\sigma} \in \Omega \mid P_{\mathrm{empirical}}(\boldsymbol{\sigma}) > 0\}$ is a set of activity patterns that appear at least once in the empirical data. The minimum probability flow method minimizes the last quantity in equation (2.15), i.e. the probability flow from activity patterns that appear in the data to the patterns that do not. Therefore, the method is not effective when N is small and $t_{\max}$ is large such that most activity patterns appear in the data. However, when N is large or $t_{\max}$ is small, this algorithm is efficient in terms of the computation time and memory space [20]. A gradient descent method on $D_{\mathrm{KL}}(P(\boldsymbol{\sigma};0) \parallel P(\boldsymbol{\sigma};\epsilon))$ is practically used for determining h and J.

(c) Accuracy indices

Fully describing an empirical distribution requires $2^N - 1$ parameters, whereas the pairwise MEM only uses $N + N(N-1)/2$ parameters. The pairwise MEM imposes that the first two moments of σ_i agree between the empirical data and the model. However, the model may be inaccurate in describing higher-order correlations in the empirical data. Most previous studies used one or both of the following two measures to quantify the accuracy with which the MEM fitted the empirical data.

The first measure is defined by

$$
\frac{I_2}{I_N} = \frac{S_1 - S_2}{S_1 - S_N}, \tag{2.16}
$$

which ranges between 0 and 1 for the maximum-likelihood estimator, and $S_k \equiv -\sum_{\boldsymbol{\sigma} \in \Omega} P_k(\boldsymbol{\sigma})$ $\log P_k(\boldsymbol{\sigma})$ is the Shannon entropy of the MEM incorporating correlations up to the kth order [17,21]. The so-called independent MEM, in which we suppress any interaction between the elements (i.e. $J_{ij} = 0$ for $i,j = 1,\ldots,N$), gives $P_1(\boldsymbol{\sigma})$. The pairwise MEM gives $P_2(\boldsymbol{\sigma})$. The empirical distribution (i.e. $P_{\mathrm{empirical}}(\boldsymbol{\sigma})$) is identical to $P_N(\boldsymbol{\sigma})$. The denominator of equation (2.16), $S_1 - S_N \equiv I_N$, is referred to as the multi-information, which quantifies the total contribution of the second or higher order correlation to the entropy of the empirical distribution. The numerator, $S_1 - S_2 \equiv I_2$, is equal to the contribution of the pairwise correlation. If $I_2/I_N = 1$, the pairwise correlation alone accounts for all the correlations present in the empirical data. If $I_2/I_N = 0$, the pairwise correlation does not deliver any information.

The second measure is defined by

$$
r = \frac{D_{\mathrm{KL}}(P_1(\boldsymbol{\sigma}) \parallel P_N(\boldsymbol{\sigma})) - D_{\mathrm{KL}}(P_2(\boldsymbol{\sigma}) \parallel P_N(\boldsymbol{\sigma}))}{D_{\mathrm{KL}}(P_1(\boldsymbol{\sigma}) \parallel P_N(\boldsymbol{\sigma}))}. \tag{2.17}
$$

Note that r also ranges between 0 and 1 for the maximum-likelihood estimator [5,12,22]. If the pairwise MEM produces a distribution closer to the empirical distribution than the independent MEM does, r is large. If the pairwise MEM and the independent MEM are similar in terms of the

proximity to the empirical distribution, we obtain $r \approx 0$. For the maximum-likelihood estimator, we obtain $I_2/I_N = r$ [5,18].

(d) Disconnectivity graph and energy landscape

Once we have estimated the pairwise MEM, we construct a dendrogram referred to as a disconnectivity graph [23], as shown in figure 1f. In the disconnectivity graph, a leaf (with a loose end open downwards) corresponds to an activity pattern σ that is a local minimum of the energy, i.e. an activity pattern whose frequency is higher than any other activity pattern in the neighbourhood of σ. The neighbourhood of σ is defined as the set of the N activity patterns that are different from σ only at one ROI. In the disconnectivity graph, the vertical position of the endpoint of the leaf represents its energy value, which specifies the frequency of appearance, with a lower position corresponding to a higher frequency. The branching structure of the disconnectivity graph describes the energy barrier between any pair of activity patterns that are local minimums. For example, to transit from local minimum α_1 to local minimum α_3 in figure 1f, the brain dynamics have to overcome the height of the energy barrier (shown by the double-headed arrow). If the barrier is high, transitions between the two activity patterns occur with a low frequency.

The disconnectivity graph is obtained by the following procedures. First, we enumerate local minimums, i.e. the activity patterns whose energy is smaller than that of all neighbours. Then, for a given pair of local minimums α and α', we consider a path connecting them, $\alpha \leftrightarrow \alpha'$, where a path is a sequence of activity patterns starting from α and ending at α' such that any two consecutive activity patterns on the path are neighbouring patterns. We denote by $E_{\max}(\alpha \leftrightarrow \alpha')$ the largest energy value among the activity patterns on path $\alpha \leftrightarrow \alpha'$. The brain dynamics on this path must climb up the hill to go through the activity pattern with energy $E_{\max}(\alpha \leftrightarrow \alpha')$ to travel between α and α'. Because a large energy value corresponds to a low frequency of the activity pattern, a large $E_{\max}(\alpha \leftrightarrow \alpha')$ value implies that the frequency of switching between α and α' along this path is low. Because various paths may connect α and α', we consider

$$\bar{E}_{\alpha\alpha'} = \min_{\alpha \leftrightarrow \alpha'} E_{\max}(\alpha \leftrightarrow \alpha'). \tag{2.18}$$

If we remove all the rarest activity patterns whose energy is equal to or larger than $\bar{E}_{\alpha\alpha'}$, α and α' are disconnected (i.e. no path connecting them exists). The energy barrier for the transition from α to α' is given by $\bar{E}_{\alpha\alpha'} - E(\alpha)$.

To calculate $\bar{E}_{\alpha\alpha'}$, we employ a Dijkstra-like method as follows. Consider the hypercube composed of 2^N activity patterns. By definition, two nodes (i.e. activity patterns) are adjacent to each other (i.e. directly connected by a link) if they are neighbouring activity patterns. Each node has degree (i.e. number of neighbours) N. Then, fix a local minimum activity pattern α and look for $\bar{E}_{\alpha\alpha'}$ for all local minimums α'. We set $\bar{E}_{\alpha\alpha} = E(\alpha)$ and $\bar{E}_{\alpha\alpha'} = E(\alpha')$ for all α' that are neighbours of α. These values are finalized and will not be changed. $\bar{E}_{\alpha\alpha'}$ for the other $2^N - N - 1$ local minimums α' are initialized to ∞. Then, we iterate the following procedures until $\bar{E}_{\alpha\alpha'}$ values for all the nodes α' are finalized. (i) For each finalized α', update $E_{\alpha\alpha''}$ for its all unfinalized neighbours α'' using

$$\bar{E}_{\alpha\alpha''}^{\text{new}} = \begin{cases} \min\{\bar{E}_{\alpha\alpha''}^{\text{old}}, \bar{E}_{\alpha\alpha'}\} & (\bar{E}_{\alpha\alpha'} \geq E(\alpha'')), \\ E(\alpha'') & (\bar{E}_{\alpha\alpha'} < E(\alpha'')). \end{cases} \tag{2.19}$$

(ii) Find α' with the smallest unfinalized $\bar{E}_{\alpha\alpha'}$ value and finalize it. (iii) Repeat steps (i) and (ii). If we carry out the entire procedure for each local minimum α, we obtain $\bar{E}_{\alpha\alpha'}$ for all pairs of local minimums.

By collecting pairs of local minimums that have the same $\bar{E}_{\alpha\alpha'}$ value, we specify a set of local minimums that should be located under the same branch. This information is sufficient for drawing the dendrogram of local minimums, i.e. the disconnectivity graph.

Each local minimum has a basin of attraction in the state space, Ω. Each activity pattern, denoted by σ, usually belongs to one of the attractive basins, which is determined as follows.

(i) Unless $\boldsymbol{\sigma}$ is a local minimum, move to the neighbouring activity pattern that has the smallest energy value. (ii) Repeat step (i) until a local minimum, denoted by $\boldsymbol{\alpha}$, is reached. We conclude that $\boldsymbol{\sigma}$ belongs to the attractive basin of $\boldsymbol{\alpha}$. (iii) Repeat steps (i) and (ii) for all the initial activity patterns $\boldsymbol{\sigma} \in \Omega$.

Using the information on the local minimums and attractive basins, the dynamics of the activity pattern are illustrated as the motion of a 'ball' on the energy landscape, as schematically shown in figure 1g as a hypothetical two-dimensional landscape. The local minimums and energy barriers in figure 1g correspond to those shown in the disconnectivity graph (figure 1f).

(e) 0/1 versus $1/-1$

We remark on two binarization schemes. In statistical physics, the pairwise MEM, or the Ising model, usually employs $\sigma_i \in \{-1, 1\}$ $(i = 1, \ldots, N)$ rather than $\tilde{\sigma}_i \in \{0, 1\}$. The former convention respects the symmetry between the two spin states and is also convenient in some analytical calculations of the model that exploit the relationship $(\sigma_i)^2 = 1$ regardless of σ_i [24,25]. For neuronal spike data, $\tilde{\sigma}_i \in \{0, 1\}$ is often used [22,26,27], whereas $\sigma_i \in \{-1, 1\}$ is also commonly used [17,18,28,29]. For fMRI data, our previous work employed $\tilde{\sigma}_i \in \{0, 1\}$ [12–15]. The use of $\tilde{\sigma}_i \in \{0, 1\}$ in representing neuronal spike trains has a rationale in being able to express the instantaneous firing rate in a simple form $\sum_{i=1}^{N} \sigma_i$ and synchronous firing of neurons by a simple multiplication [30]. For example, three neurons simultaneously fire if and only if $\sigma_1 \sigma_2 \sigma_3 = 1$. It should also be noted that the iterative scaling algorithm for maximizing the likelihood (§2b(i)) does not generally work for $\sigma_i \in \{-1, 1\}$ because $\langle \sigma_i \rangle_{\text{empirical}}$ and $\langle \sigma_i \rangle_{\text{model}}$, the logarithm of whose ratio is used in the algorithm, may have opposite signs.

The energy in the case of $\tilde{\sigma}_i \in \{0, 1\}$ is defined as

$$\tilde{E}(\tilde{\boldsymbol{\sigma}} \mid \tilde{\boldsymbol{h}}, \tilde{\mathbf{J}}) = -\sum_{i=1}^{N} \tilde{h}_i \tilde{\sigma}_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \tilde{J}_{ij} \tilde{\sigma}_i \tilde{\sigma}_j. \tag{2.20}$$

Mathematically, the two representations are equivalent to the one-to-one relationship, $2\tilde{\sigma}_i - 1 = \sigma_i$ $(i = 1, \ldots, N)$, which results in

$$\tilde{h}_i = 2h_i - 2 \sum_{\substack{j=1 \\ j \neq i}}^{N} J_{ij} \tag{2.21}$$

and

$$\tilde{J}_{ij} = 4J_{ij}. \tag{2.22}$$

3. Results

(a) Accuracy of the three methods

We applied the three methods to estimate the pairwise MEM to resting-state fMRI signals recorded from two healthy adult individuals in the Human Connectome Project. We extracted ROIs from three brain systems, i.e. default mode network (DMN, $N_{\text{ROI}} = 12$), fronto-parietal network (FPN, $N_{\text{ROI}} = 11$) and cingulo-opercular network (CON, $N_{\text{ROI}} = 7$), using the ROIs whose coordinates were identified previously [31]. We had $t_{\max} = 9560$ volumes in total.

The estimated parameter values are compared between likelihood maximization and pseudo-likelihood maximization in figure 2a–f. For all the networks, the results obtained by the pseudo-likelihood maximization are close to those obtained by the likelihood maximization, in particular for $\mathbf{J}$. The results obtained by the likelihood maximization and those obtained by the minimum probability flow are compared in figure 2g–j for the DMN and FPN. We did not apply the minimum probability flow method to the CON because all of the 2^8 activity patterns appeared at least once, i.e. $\mathcal{D} = \Omega$, which made the right-hand side of equation (2.15) zero. The
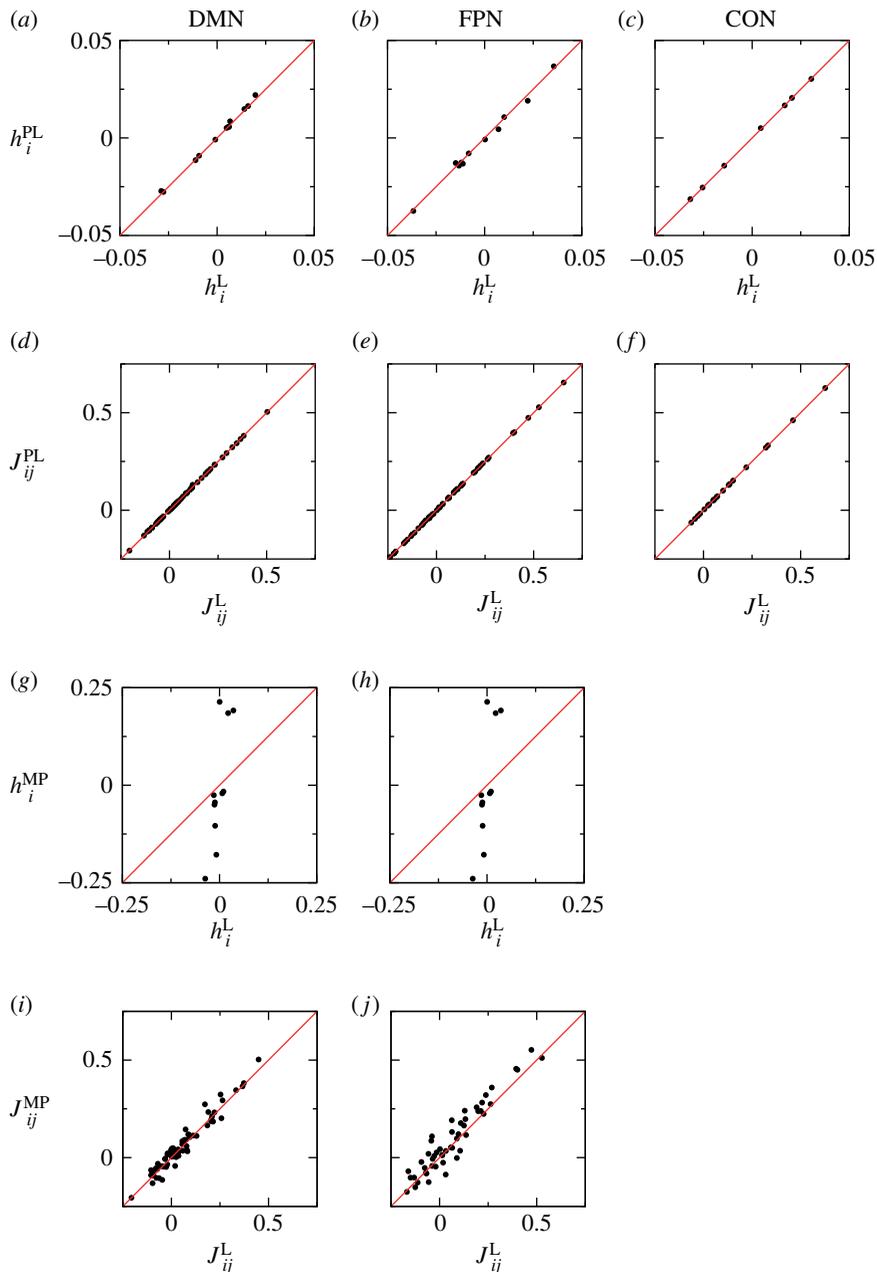
9

rsta.royalsocietypublishing.org   *Phil. Trans. R. Soc. A* **375**: 20160287

Figure 2. Estimated parameter values compared between the different methods. (*a*–*f*) Comparison between the likelihood maximization and the pseudo-likelihood estimation. (*g*–*j*) Comparison between the likelihood maximization and the minimum probability flow method. (*a,d,g,i*) DMN, (*b,e,h,j*) FPN and (*c,f*) CON. Superscripts L, PL and MP stand for likelihood maximization, pseudo-likelihood maximization and the minimum probability flow, respectively. (Online version in colour.)

figure indicates that the estimation by the minimum probability flow deviates from that by the likelihood maximization more than the estimation by the pseudo-likelihood maximization does, in particular for h. The two measures of the accuracy indices are shown in table 1 for each network and estimation method. The two indices took the same value in the case of the likelihood maximization [5,18]. In the case of the pseudo-likelihood maximization, the two accuracy indices were slightly different from each other, and both took approximately the same values as those derived from the maximum likelihood. In the case of the minimum probability flow, r was
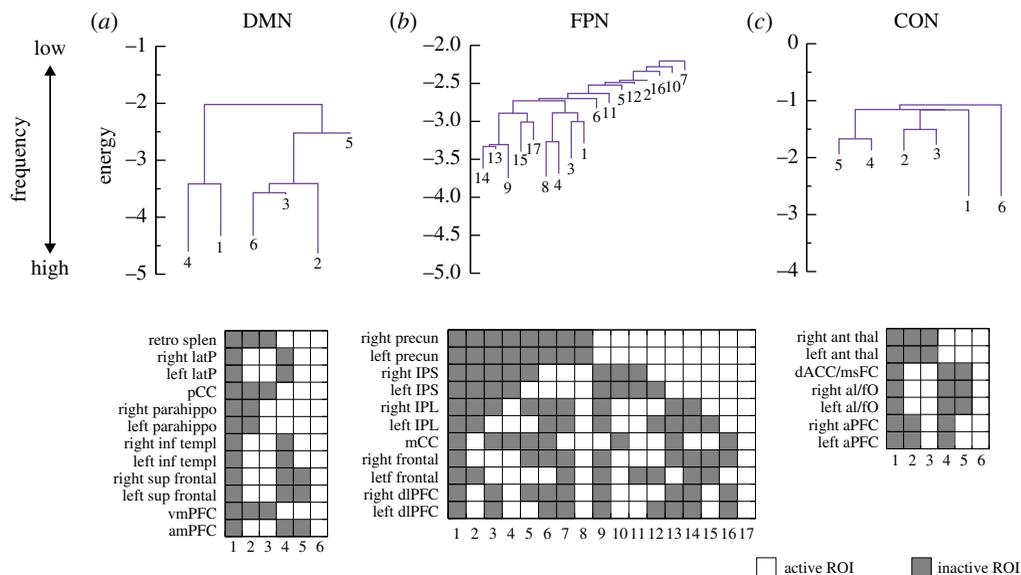
Figure 3. Disconnectivity graphs for (*a*) DMN, (*b*) FPN and (*c*) CON. The activity pattern at each local minimum is also shown. Retro splen: retro splenial cortex, latP: lateral parietal cortex, pCC: posterior cingulate cortex, parahippo: parahippocampal cortex, inf templ: inferior temporal cortex, sup frontal: superior frontal cortex, vmPFC: ventromedial prefrontal cortex, amPFC: anteromedial prefrontal cortex, precun: precuneus, IPS: intraparietal sulcus, IPL: inferior parietal lobule, mCC: mid-cingulate cortex, frontal: lateral frontal cortex, dlPFC: dorsolateral prefrontal cortex, ant thal: anterior thalamus, dACC/msFC: dorsal anterior cingulate cortex/medial superior frontal cortex, aI/fO: anterior insula/frontal operculum, aPFC: anterior prefrontal cortex. (Online version in colour.)

Table 1. Accuracy of fitting for each network and estimation algorithm.

	DMN		FPN		CON	
	r	I_2/I_N	r	I_2/I_N	r	I_2/I_N
likelihood maximization	0.6921	0.6921	0.7830	0.7830	0.9744	0.9744
pseudo-likelihood maximization	0.6921	0.6972	0.7830	0.7853	0.9745	0.9744
minimum probability flow	0.6480	0.7437	0.6124	1.2295	—	—

substantially smaller than the values for the likelihood or pseudo-likelihood maximization. By contrast, I_2/I_N exceeded unity because $S_1 > S_N > S_2$ for the minimum probability flow method.

(b) Disconnectivity graphs

Figure 3 shows the disconnectivity graph of the DMN, FPN and CON, calculated for the parameter values estimated by likelihood maximization. The two synchronized activity patterns, i.e. the activity patterns with all ROIs being active or inactive, were local minimums. The FPN had much more local minimums than the DMN and CON did. Although the present results are opposite to our previous results using a different dataset [14], the reason for the discrepancy is unclear.

(c) Effects of the data length

Our experiences suggest that, as the number of ROIs, N, increases, the pairwise MEM seems to demand a large amount of data to realize a high accuracy. If we use N ROIs, there are 2^N
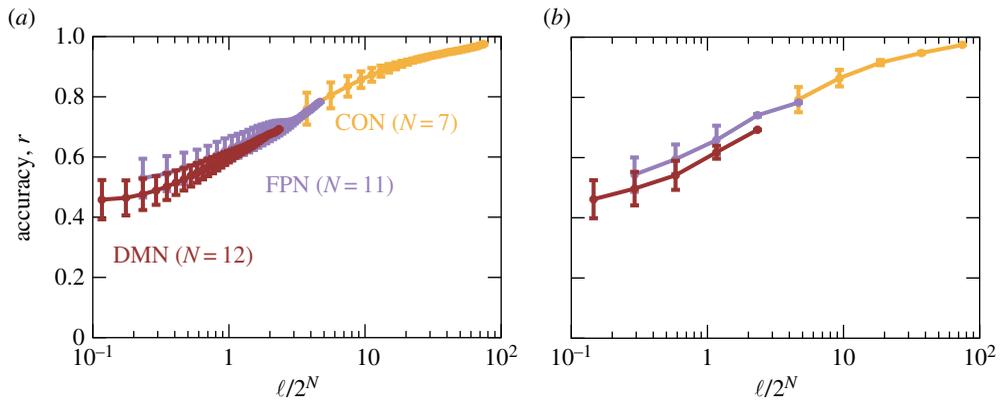
Figure 4. Accuracy of the fit of the pairwise MEM as a function of the data length, ℓ, normalized by the number of possible activity patterns, 2^N. We obtained samples of length ℓ by (*a*) overlapping sliding time windows and (*b*) non-overlapping time windows. Error bars represent the standard deviation. (Online version in colour.)

possible activity patterns. Therefore, as we increase N, it is progressively more likely that many of the activity patterns are unvisited. However, the MEM assigns a positive probability to such an unvisited pattern. Even if an activity pattern σ is realized by the empirical data a few times, the empirical distribution, $P_{\text{empirical}}(\sigma)$, would not be reliable because it is evaluated only based on a few visits to σ (divided by $t_{\max}$). If $t_{\max}$ is much larger and σ is visited proportionally many times, then we would be able to estimate $P_{\text{empirical}}(\sigma)$ more accurately. This exercise led us to hypothesize that the accuracy scales as a function of $t_{\max}/2^N$.

To examine this point, we carried out likelihood maximization on the fMRI data of varying length ℓ ($t_{\max}/20 \leq \ell \leq t_{\max}$) and calculated r (which coincides with I_2/I_N for the maximum-likelihood estimator). For a given ℓ, we calculated r for each of the $t_{\max} - \ell$ datasets of length ℓ, i.e. $\{\sigma(1), \ldots, \sigma(\ell)\}$, $\{\sigma(2), \ldots, \sigma(\ell + 1)\}$, $\ldots$, $\{\sigma(t_{\max} - \ell + 1), \ldots, \sigma(t_{\max})\}$. The average and standard deviation of r as a function of $\ell/2^N$ are shown in figure 4a for the DMN, FPN and CON. As expected, the accuracy improved as ℓ increased. The results for the DMN, FPN and CON roughly collapsed on a single curve. The figure suggests that, to achieve an accuracy of 0.8 and 0.9, each activity pattern should be visited ≈ 5 and ≈ 16 times on average, respectively.

Because the aforementioned sampling method used overlapping time windows to make different samples strongly depend on each other, we carried out the same test by dividing the entire time series $\{\sigma(1), \ldots, \sigma(t_{\max})\}$ into two halves of length $\ell = t_{\max}/2$, four quarters of length $\ell = t_{\max}/4$, eight non-overlapping samples of length $\ell = t_{\max}/8$ and so forth. The results (figure 4b) were similar to those in the case of overlapping time windows (figure 4a).

4. Discussion

We explained a set of computational methods to estimate the pairwise MEM and energy landscapes from resting-state fMRI data. Novel components, as compared with our previous methods [12–15], were the pseudo-likelihood maximization, the minimum probability flow and a variant of the Dijkstra method to calculate the disconnectivity graph. We applied the methods to fMRI data collected from healthy participants and assessed the amount of data needed to secure a sufficient accuracy of fit.

The present results suggest that the current method is admittedly demanding in terms of the amount of data, although the results should be corroborated with different datasets. In the application of the pairwise MEM to neuronal spike trains, the data length does not seem to pose a severe limit if the network size, N, is comparable to those in this study. This is because one typically uses a high time resolution to ensure that there are no multiple spikes within a time window (e.g. 2 ms [28], 10 ms [22], 20 ms [17,18,27]). Then, the number of data points, $t_{\max}$,

is typically much larger than in typical fMRI experiments. In fMRI experiments, the interval between two measurements, called the repetition time (TR), is typically 2–4 s, and a participant in the resting state (or a particular task condition) can be typically scanned for 5–15 min. Then, we would have $t_{max} = 75$–450, with which we can reliably estimate the pairwise MEM model up to $N \approx 5$ (figure 4), which is small. If we pool fMRI data from 10 participants belonging to the same group to estimate one MEM, we would have $t_{max} = 750$–4500, accommodating $N \approx 8$. This is an important limitation of our approach. Currently we cannot apply the method to relatively large brain systems (i.e. those with a larger number of ROIs), let alone voxel-based data.

We demonstrated the methods with fMRI data obtained from healthy participants. The same methods can be applied to different conditions of human participants including the case of medical applications, the topic of the present theme issue. Various neuropsychiatric disorders have been suggested to have dynamical footprints in the brain [1–4]. Altered dynamics in the brain at various spatial and time scales may result in deformation of energy landscapes as compared with healthy controls.

5. Material and methods

(a) Data and participants

We used resting-state fMRI data publicly shared in the Human Connectome Project (acquisition Q10 in release S900 of the WU-Minn HCP data) [32]. The data were collected using a 3T MRI (Skyra, Siemens) with an echo planar imaging (EPI) sequence (TR, 0.72 s; TE, 33.1 ms; 72 slices; 2.0 mm isotropic; field of view, 208×180 mm) and T1-weighted sequence (TR, 2.4 s; TE, 2.14 ms; 0.7 mm isotropic; field of view, 224×224 mm). The EPI data were recorded in four runs ($\approx$15 min run^{-1}) while participants were instructed to relax while looking at a fixed cross mark on a dark screen.

We used such EPI and T1 images recorded from two adult participants (one female; 22–25 years old), because the amount of the data was sufficiently large for the current analysis.

(b) Preprocessing and extraction of region of interest data

We preprocessed the EPI data in essentially the same manner as the conventional methods that we previously used for resting-state fMRI data [12,33,34] with SPM12 (www.fil.ucl.ac.uk/spm). Briefly, after discarding the first five images in each run, we conducted realignment, unwarping, slice timing correction, normalization to the standard template (ICBM 152) and spatial smoothing (full-width at half maximum = 8 mm). Afterwards, we removed the effects of head motion, white matter signals and cerebrospinal fluid signals by a general linear model. Finally, we performed temporal band-pass filtering (0.01–0.1 Hz) and obtained resting-state whole-brain data.

We then extracted a time series of fMRI signals from each ROI. The ROIs were defined as 4 mm spheres around their centre whose coordinates were determined in a previous study [31]. The signals at each ROI were those averaged over the sphere. In total, we obtained time-series data of length $t_{max} = 9560$ at 30 ROIs (12 in the DMN, 11 in the FPN and 7 in the CON).

References

1. Rolls ET, Deco G. 2011 A computational neuroscience approach to schizophrenia and its onset. *Neurosci. Biobehav. Rev.* **35**, 1644–1653. (doi:10.1016/j.neubiorev.2010.09.001)

2. Uhlhaas PJ, Singer W. 2012 Neuronal dynamics and neuropsychiatric disorders: toward a translational paradigm for dysfunctional large-scale networks. *Neuron* **75**, 963–980. (doi:10.1016/j.neuron.2012.09.004)

3. Kopell NJ, Gritton HJ, Whittington MA, Kramer MA. 2014 Beyond the connectome: the dynome. *Neuron* **83**, 1319–1328. (doi:10.1016/j.neuron.2014.08.016)

4. Wang XJ, Krystal JH. 2014 Computational psychiatry. *Neuron* **84**, 638–654. (doi:10.1016/j.neuron.2014.10.018)

5. Yeh FC, Tang A, Hobbs JP, Hottowy P, Dabrowski W, Sher A, Litke A, Beggs JM. 2010 Maximum entropy approaches to living neural networks. *Entropy* **12**, 89–106. (doi:10.3390/e12010089)

6. Das T, Abeyasinghe PM, Crone JS, Sosnowski A, Laureys S, Owen AM, Soddu A. 2014 Highlighting the structure-function relationship of the brain with the Ising model and graph theory. *BioMed Res. Int.* **2014**, 237898. (doi:10.1155/2014/237898)

7. Schneidman E. 2016 Towards the design principles of neural population codes. *Curr. Opin. Neurobiol.* **37**, 133–140. (doi:10.1016/j.conb.2016.03.001)

8. Fraiman D, Balenzuela P, Foss J, Chialvo DR. 2009 Ising-like dynamics in large-scale functional brain networks. *Phys. Rev. E* **79**, 061922. (doi:10.1103/PhysRevE.79.061922)

9. Chialvo DR. 2010 Emergent complex neural dynamics. *Nat. Phys.* **6**, 744–750. (doi:10.1038/nphys1803)

10. Hudetz AG, Humphries CJ, Binder JR. 2014 Spin-glass model predicts metastable brain states that diminish in anesthesia. *Front. Syst. Neurosci.* **8**, 234. (doi:10.3389/fnsys.2014.00234)

11. Marinazzo D, Pellicoro M, Wu G, Angelini L, Cortés JM, Stramaglia S. 2014 Information transfer and criticality in the Ising model on the human connectome. *PLoS ONE* **9**, e93616. (doi:10.1371/journal.pone.0093616)

12. Watanabe T, Hirose S, Wada H, Imai Y, Machida T, Shirouzu I, Konishi S, Miyashita Y, Masuda N. 2013 A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nat. Commun.* **4**, 1370. (doi:10.1038/ncomms2388)

13. Watanabe T, Kan S, Koike T, Misaki M, Konishi S, Miyauchi S, Miyahsita Y, Masuda N. 2014 Network-dependent modulation of brain activity during sleep. *Neuroimage* **98**, 1–10. (doi:10.1016/j.neuroimage.2014.04.079)

14. Watanabe T, Hirose S, Wada H, Imai Y, Machida T, Shirouzu I, Konishi S, Miyashita Y, Masuda N. 2014 Energy landscapes of resting-state brain networks. *Front. Neuroinform.* **8**, 12. (doi:10.3389/fninf.2014.00012)

15. Watanabe T, Masuda N, Megumi F, Kanai R, Rees G. 2014 Energy landscape and dynamics of brain activity during human bistable perception. *Nat. Commun.* **5**, 4765. (doi:10.1038/ncomms5765)

16. Darroch JN, Ratcliff D. 1972 Generalized iterative scaling for log-linear models. *Ann. Math. Stat.* **43**, 1470–1480.

17. Schneidman E, Berry MJ, Segev R, Bialek W. 2006 Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012. (doi:10.1038/nature04701)

18. Tang A *et al.* 2008 A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J. Neurosci.* **28**, 505–518. (doi:10.1523/JNEUROSCI.3359-07.2008)

19. Besag J. 1975 Statistical analysis of non-lattice data. *J. R. Stat. Soc. D* **24**, 179–195. (doi:10.2307/2987782)

20. Sohl-Dickstein J, Battaglino PB, DeWeese MR. 2011 New method for parameter estimation in probabilistic models: minimum probability flow. *Phys. Rev. Lett.* **107**, 220601. (doi:10.1103/PhysRevLett.107.220601)
21. Schneidman E, Still S, Berry MJ, Bialek W. 2003 Network information and connected correlations. *Phys. Rev. Lett.* **91**, 238701. (doi:10.1103/PhysRevLett.91.238701)
22. Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, Sher A, Litke AM, Chichilnisky EJ. 2006 The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* **26**, 8254–8266. (doi:10.1523/JNEUROSCI.1282-06.2006)
23. Becker OM, Karplus M. 1997 The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. *J. Chem. Phys.* **106**, 1495–1517. (doi:10.1063/1.473299)
24. Mézard M, Parisi G, Virasoro MA 1987 *Spin glasses and beyond*. Singapore: World Scientific.
25. Nishimori N. 2001 *Statistical physics of spin glasses and information processing—an introduction*. Oxford, UK: Oxford University Press.
26. Roudi Y, Nirenberg S, Latham PE. 2009 Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't. *PLoS. Comput. Biol.* **5**, e1000380. (doi:10.1371/journal.pcbi.1000380)
27. Ganmor E, Segev R, Schneidman E. 2011 The architecture of functional interaction networks in the retina. *J. Neurosci.* **31**, 3044–3054. (doi:10.1523/JNEUROSCI.3682-10.2011)
28. Yu S, Huang D, Singer W, Nikolić D. 2008 A small world of neuronal synchrony. *Cereb. Cortex* **18**, 2891–2901. (doi:10.1093/cercor/bhn047)
29. Roudi Y, Tyrcha J, Hertz J. 2009 Ising model for neural data: model quality and approximate methods for extracting functional connectivity. *Phys. Rev. E* **79**, 051915. (doi:10.1103/PhysRevE.79.051915)
30. Amari S, Nakahara H, Wu S, Sakai Y. 2003 Synchronous firing and higher-order interactions in neuron pool. *Neural. Comput.* **15**, 127–142. (doi:10.1162/089976603321043720)
31. Fair DA, Cohen AL, Power JD, Dosenbach NUF, Church JA, Miezin FM, Schlaggar BL, Petersen SE. 2009 Functional brain networks develop from a 'local to distributed' organization. *PLoS. Comput. Biol.* **5**, e1000381. (doi:10.1371/journal.pcbi.1000381)
32. Van Essen DC *et al.* 2012 The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222–2231. (doi:10.1016/j.neuroimage.2012.02.018)
33. Watanabe T *et al.* 2015 Clinical and neural effects of six-week administration of oxytocin on core symptoms of autism. *Brain* **138**, 3400–3412. (doi:10.1093/brain/awv249)
34. Watanabe T *et al.* 2015 Effects of rTMS of pre-supplementary motor area on fronto basal ganglia network activity during stop-signal task. *J. Neurosci.* **35**, 4813–4823. (doi:10.1523/JNEUROSCI.3761-14.2015)