Characterization of network structure in stereoEEG data using consensus-based partial coherence

Marije Ter Wal a,*,1, Pasquale Cardellicchio b, Giorgio LoRusso c, Veronica Pelliccia c, Pietro Avanzini b, Guy A. Orban b, Paul H.E. Tiesinga a

a Donders Institute, Radboud University, Nijmegen, The Netherlands
b Department of Neuroscience, University of Parma, Parma, Italy
c Claudio Munari Center for Epilepsy Surgery, Ospedale Niguarda Ca’ Granda, Milano, Italy

ARTICLE INFO

Keywords:
Coherence
Partial coherence
Consensus
Connectivity

ABSTRACT

Coherence is a widely used measure to determine the frequency-resolved functional connectivity between pairs of recording sites, but this measure is confounded by shared inputs to the pair. To remove shared inputs, the ‘partial coherence’ can be computed by conditioning the spectral matrices of the pair on all other recorded channels, which involves the calculation of a matrix (pseudo-) inverse. It has so far remained a challenge to use the time-resolved partial coherence to analyze intracranial recordings with a large number of recording sites. For instance, calculating the partial coherence using a pseudoinverse method produces a high number of false positives when it is applied to a large number of channels.

To address this challenge, we developed a new method that randomly aggregated channels into a smaller number of effective channels on which the calculation of partial coherence was based. We obtained a ‘consensus’ partial coherence (cPCOH) by repeating this approach for several random aggregations of channels (permutations) and only accepting those activations in time and frequency with a high enough consensus.

Using model data we show that the cPCOH method effectively filters out the effect of shared inputs and performs substantially better than the pseudo-inverse. We successfully applied the cPCOH procedure to human stereotactic EEG data and demonstrated three key advantages of this method relative to alternative procedures. First, it reduces the number of false positives relative to the pseudo-inverse method. Second, it allows for titration of the amount of false positives relative to the false negatives by adjusting the consensus threshold, thus allowing the data-analyst to prioritize one over the other to meet specific analysis demands. Third, it substantially reduced the number of identified interactions compared to coherence, providing a sparser network of connections from which clear spatial patterns emerged. These patterns can serve as a starting point of further analyses that provide insight into network dynamics during cognitive processes. These advantages likely generalize to other modalities in which shared inputs introduce confounds, such as electroencephalography (EEG) and magnetoencephalography (MEG).

Introduction

A central task in neuroimaging is the identification of connections in the brain, either structural, i.e. mediated by direct axonal connections, or functional, i.e. related to the task, context or brain state. Such connections form the distributed network that underlies all complex brain functions, from hierarchical processing of inputs in the sensory cortices to memory formation, decision making and other cognitive functions in the parietal and frontal cortices. Functional connectivity, unlike structural connectivity, is correlational in nature, and is identified as a statistical dependence between time courses of neural activity recorded in the brain. In humans, functional connections have been identified using fMRI, MEG/EEG, as well as invasive ECoG and stereotactic EEG (stereoEEG or sEEG) recordings. Analysis methods include correlation-,
coherence- (Palva and Palva, 2012; Pereda et al., 2005; Salvador et al., 2005) and information theory-based (Pereda et al., 2005) data-driven approaches as well as model-dependent approaches such as Granger Causality (Geweke, 1982; Seth et al., 2015; Vinck et al., 2015), for an overview see (Bastos and Schoffelen, 2016; Sakkalis, 2011). For electrophysiological data, functional connectivity is usually defined as synchronization between the oscillatory activities of a pair of recording sites. Functional connections are then inferred from a co-modulation of power (i.e. coherence) and/or phase in the studied frequency band.

Coherence analysis (COH) identifies linear interactions and can reflect both direct and indirect connections between two cortical areas. The precise link between coherence and the functional relevance of the underlying connection, be it selection of information (Bosman et al., 2012) or information transfer (Fries, 2015), network formation or changes (Colgin et al., 2009), or otherwise, is still a topic of further study. Despite these interpretational limitations, coherence is a popular analysis technique, due to its wide applicability and computational simplicity. Furthermore, unlike other approaches, coherence allows for the identification of several co-existing networks in different frequency bands, and can track them over time.

Due to its correlational nature, coherence analysis is prone to spurious results caused by common sources of activity (Bastos and Schoffelen, 2016). A common source can arise from volume conducted activity from a nearby region, which is of specific concern for EEG, and to a much lesser extent, MEG recordings. But common sources can also have a neuronal basis, and hence also affect recording techniques that are not prone to volume conduction, such as stereoEEG (Caruana et al., 2017): a pair of recording sites driven by activity from the same, third location, prone to volume conduction, such as stereoEEG (Caruana et al., 2017); a pair of recording sites driven by activity from the same, third location, will produce a high coherence, despite having no structural or functional neuronal basis, and hence also affect recording techniques that are not validated the partial coherence for datasets larger than approximately 10 channels (Fig. 1B). The method can still be used by applying a pseudoinverse algorithm, but results are likely to be less reliable. Indeed, we report here that the pseudoinverse-based partial coherence (PINV) produces many spurious responses (false positives) and removes legitimate responses, even when averaging across all trials in the task condition, rather than considering single trials.

To improve the computation of partial coherence for large datasets, we introduce a new consensus-based implementation of partial coherence (cPCOH). This method reduces the dataset size by averaging channels within randomly selected groups and computing partial coherence based on these averages. By repeating this procedure across many random groupings, a consensus across permutations can be established.

We assess the performance of the consensus-based partial coherence in a model dataset, as well as on sEEG data from two patients. We find that cPCOH substantially outperforms a pseudoinverse based partial coherence method in both the model and the human data: it introduces fewer false positives and removes fewer direct connections (fewer false negatives). In addition, we report that partial coherence has a substantial impact on the findings from the sEEG dataset: cPCOH filters roughly three quarters of coherence patterns found, and produces more precise, more restricted functional networks.
Materials & methods

**Human stereoEEG dataset**

The consensus-based partial coherence method was developed for task-related connectivity analysis of human stereoEEG (sEEG) data. sEEG is an invasive technique that uses depth electrodes, which are stereotactically inserted into the cerebral hemispheres. It is used in epilepsy patients for identification of the epileptic focus/foci (Cossu et al., 2005; Munari et al., 1994).

**Patients and implantation**

Here, we present data from two patients (P102 and P103, 1 female) recorded in the Claudio Munari Center for Epilepsy Surgery of the Ospedale Niguarda-Ca’ Granda in Milano, Italy. Sixteen (P102) and thirteen (P103) cylindrical electrodes (DIXI Medical, Besancon, France), respectively, were implanted in the frontal and temporal lobes of the right hemisphere. The electrodes had a diameter of 0.8 mm and recording leads of 1.5 mm long, spaced 2 mm apart. The electrical characteristics and spacing of the leads are expected to result in local signals and minimal volume conduction, as was previously shown in a related dataset (Caruana et al., 2017). The data used here followed a similar pattern (Fig. S2A).

Patients were informed of the electrode implantation and sEEG recordings, and gave their written informed consent to participate in the study according to the Declaration of Helsinki. Experimental procedures were approved by the Ethical-Scientific Committee of the Ospedale Niguarda-Ca’ Granda. Implantation sites were selected purely on clinical grounds, on the basis of seizure semiology, scalp-EEG and neuroimaging studies. No seizures were observed during the 24 h prior to the recording sessions. Electrode placement was confirmed by post operation CT-scans co-registered to the pre-operative MR scans using FLIRT (Jenkinson and Smith, 2001). Recording leads were recovered by thresholding the CT images and localized relative to the cortical ribbon using in-house software (Avanzini et al., 2016). For the analysis presented here, leads in white matter were excluded. For patient P102, 119 leads in the right hemisphere were kept, which are shown on the gray matter flat map in Fig. 2A. For patient P103, 115 leads in the right hemisphere remained.

**Task**

The patients performed a rule-switching task adapted from Buschman et al. (2012) in macaques. In the adjusted version of the task, subjects were presented with an elongated visual stimulus, comprised of a sequence of M’s and W’s, and were asked to judge either its color (red or blue) or orientation (horizontal or vertical), by pressing the left or right button as shown in Fig. 2B. Each trial started with the appearance of a fixation dot in the center of the screen, followed 100 ms later by the appearance of a dotted boundary near the frame’s edge, serving as a cue to indicate the currently active rule (Fig. 2C). Stimulus onset time was randomly chosen for each trial from 18 predefined values, ranging from 300 to 600 ms after trial start. Subjects were instructed to respond as quickly as possible. The trial timed out if the subject did not respond.

---

**Fig. 2.** A: Flat map of the right hemisphere of P102 with each of the 119 gray matter recording sites indicated by black dots. The red lines indicate pairs connecting the Superior Temporal Gyrus (STG) and the lateral Frontal Gyrus (lFG, solid) and STG and ventral Premotor Cortex (vPMC, dotted). Data from these pairs are shown in D of this figure and in Fig. 7; B: The desired outputs (left or right button press) for the task rules and for the different stimulus characteristics; in white triangles the desired response for the color rule, in black triangles the response for the orientation rule. C: Trial setup, see main text for description; D: Illustration of used dataset: Examples of raw sEEG traces and average spectrograms (Z-scored relative to time-shuffled reference) for the pairs indicated by the red lines in A.
within 1500 ms of stimulus onset. Since each stimulus included color and orientation (see Fig. 2B&C), the responses could be either congruent for the two rules (30% congruent trials) or not (70% incongruent trials). The rule switched between the two alternatives after a minimum of 20 trials and a maximum of 40 trials, chosen randomly, adding up to 400 presented trials for each rule.

**Data acquisition and preprocessing**

Data were recorded at a sampling frequency of 1 kHz in two continuous blocks of 400 trials. The data were zero-phase band-pass filtered with a 6th order Butterworth filter with cut-off frequencies at 3 and 300 Hz. In addition, 50 Hz line noise was removed by applying a notch filter with a bandwidth of 1 Hz at -3 dB. Recordings were inspected by clinicians and any pathological channels removed. The signals of the remaining channels were divided into trials. Trials with an incorrect response or no response were excluded from analysis. The first two trials after a rule switch were also excluded. For subject P102, this resulted in 381 and 382 remaining trials, for the color and orientation rules, respectively, and for subject P103 384 (color) and 383 (orientation) trials remained.

**Model dataset**

To test and compare the performance of coherence, pseudoinverse-based partial coherence and consensus-based partial coherence we used a model that resembled the recorded signals and task structure of the stereoEEG dataset. The model data were generated by N noisy oscillators (with N between 20 and 110 channels), modeled together as a Vector Autoregressive (VAR) process, with random frequencies, which were driven by noisy inputs and connected by a maximum of $N_{\text{con}} \cdot N$ random connections, where $N_{\text{con}}$ is the maximum number of connections per channel (Fig. 3). $N_{\text{con}}$ was varied between 5 (sparse network) and 25 (dense network) for a network of 110 channels. Details about the signals in the network are provided in section 1.2 of the Supplemental Information. As in the recorded dataset (see below), a reference dataset (surrogates) was needed for assessment of significance. In the model, the reference dataset was created by rerunning the generative model with a different seed for the random number generator. Similar to the human data, the test dataset and reference dataset consisted of 400 trials.

The coherence, pseudo-inverse and consensus-based partial coherence were computed for one pair of channels in the data: channel 1 and channel 2. To introduce non-stationarity in the network, for this pair only, an ‘event’ was created at $t = 0$: for $t < 0$, the connection between channel 1 and channel 2 had strength $c_{12}/2$, while for $t > 0$, the connection strength increased to $c_{12}$. The reference data did not have an event, and connection strength between channel 1 and 2 remained $c_{12}/2$ throughout the trial. Both members of the pair of interest received shared input from a third channel, with connection strength $c_3$. This third channel also projected to $N_{\text{shared}}$ other channels in the network with strength $c_3$. Connection between channels 4 – $N$ in the network had strength $c_3/N_{\text{con}}$, with the scaling by $N_{\text{con}}$ required to ensure stability of the model.

To determine the performance of the conditioning methods, ‘ground truth’ data were computed by breaking the shared connection from channel 3 to channel 1 and 2 (Fig. 3C). Since channel 1 and channel 2 did not project to other channels, the removal of the shared connection from channel 3 did not affect the dynamics of other channels in the network. All other connections remained intact, and the inputs used to generate the ground truth data were identical to those used for the test and reference data. The coherence between channel 1 and 2 of the ground

---

**Fig. 3.** A: Schematic representation of the model network, with circles representing VAR-variables and lines indicating connections. $c_{12}$ is the connection strength between channel 1 and 2, $c_3$ represents the connection strength for the shared connection, i.e. from channel 3 to channels 1, 2 and others; B: A reference dataset, used for computing significance, was generated by reconnecting the VAR-processes from panel A with different random connections, and reducing the connection strength between channel 1 and 2 to $c_{12}/2$; C: A ground truth dataset was generated to assess the performance of cPCOH and PINV, by removing the shared connections emanating from channel 3; D: Example of coherence (left), partial coherence (middle) and ground truth coherence (right) for one instance of the model, with a 40 Hz direct coherence due to the activity in channel 1 (dotted white box) and a 60 Hz shared input from channel 3 (solid white box). Two cPCOH permutations are shown, with one removing and one keeping the shared input. The consensus between permutations was 1 for the 40 Hz direct input, and around 0.7 for the 60 Hz shared input, hence the consensus-based partial coherence method could remove the shared input by thresholding the consensus at 0.7.
truth dataset was considered the desired output of the conditioning methods and was therefore used to score the performance of the pseudoinverse-based and consensus-based methods while the parameters $N$, $N_{\text{trials}}$, $C_{12}$, $C_{3}$, $C_{N}$, and the signal to noise ratio (SNR, see SI section 3.1) were varied. Fig. 3D shows example coherence, consensus-based partial coherence and ground truth coherence for one instance of the model (the corresponding traces and spectra are found in Supplemental Fig. S1).

Wavelet coherence and partial coherence

The trial data were transformed to the time- and scale- (i.e. pseudo-frequency) domain using a Morlet wavelet transform, with the scales logarithmically covering the 20–120 Hz frequency band and the wavelet center shifted in time increments of 5 ms. We used a relatively narrow wavelet width of 3 periods, as Augmented Dickey-Fuller tests suggested that stationarity of the data was not guaranteed for lags exceeding 100 ms (Fig. S2B), i.e. 2 periods at 20 Hz. Wavelet width had no marked effect on the outcome of the cPCOH analysis (Fig. S2C). Frequencies below 20 Hz were not considered, because the pre-cue baseline was too short to obtain meaningful statistics for lower frequencies.

The wavelet spectrograms were used to compute the wavelet coherence (Lachaux et al., 2002) between all pairs of channels in the dataset. The wavelet coherence was computed as:

$$W_{\text{coh}}(t, s) = \frac{\left| W_{ij}(t, s) \right|}{\sqrt{W_{ii}(t, s) W_{jj}(t, s)}}$$

(1)

where $W_{ij}$ is the trial-averaged cross-spectrum for channel $i \neq j$ and $W_{ii}$ is the trial-averaged auto-spectrum, for every 'pixel' in time ($t$) and scale ($s$). Here, ‘trial-averaged’ means that the cross- and auto-spectra were first computed for every trial individually, and subsequently averaged.

The initial stages and general expression for the partial coherence analysis are identical to the computation of coherence. But instead of the averaged auto- and cross-spectra, the conditioned auto- and cross-spectra are used in equation (1). The conditioned auto- and cross-spectra were obtained as described in Rosenberg et al. (1998):

$$C_{\text{cond}}(i, j) = C_{ij} - C_{ik} C_{kj}^{-1} C_{ij}$$

(2)

Here $C$ is the full and symmetric matrix of all cross-spectral densities. In other words, it combines all cross-spectra ($i \neq j$) and auto-spectra ($i = j$), with the indices $i, j = 1 \ldots N$ indicating the coordinates of the pair's channels in $C$, such that $C_{ij} = W_{ij}(t, s)$. Indices $k \ldots$ indicate the other $N-2$ channels the pair is conditioned on. For readability, the argument $(t, s)$ was omitted in the above equation, but keep in mind that this operation is performed for each time- and scale pixel separately. The second term on the right-hand side of equation (2) is the conditioning term. It assesses the linear influence of all other channels on the spectrum of pair $i, j$ and subtracts it.

Fig. 4. A: Schematic representation of the consensus-based partial coherence method. See main text for explanation; B: Random conditioning involves the random grouping of all channels that are not in the pair under study into a small number of groups. Cross- and auto-spectra are averaged within those groups and used for conditioning. The random grouping and conditioning is repeated $P$ (here: 50) times. The agreement between these repetitions (permutations) is called the 'consensus'. Only pixels with a high consensus are accepted and pixels with low consensus are masked; C: Illustration of the trial shuffling procedure used to generate the reference data. Trials for the first channel in the pair are maintained, while for the second channel data from the other rule are used, from which trials are randomly selected.
The partial cross-spectrum $C_{\text{cond}}^{ij}$ and partial auto-spectra $C_{\text{cond}}^{ij}$ and $C_{\text{cond}}^{ii}$ can subsequently replace the respective $W$ terms in equation (1), giving the partial coherence of the pair under study. To obtain the partial cross- and auto-spectra, the $C_{\text{cond}}^{ij}$ matrix needs to be inverted. For the dataset obtained from the sEEG recordings this matrix was usually ill-conditioned, and therefore the Matlab pseudoinverse algorithm pinv was used to invert the matrix, which uses a singular value decomposition to find the inverse (for more details see Press et al., 2007).

**Consensus-based partial coherence**

Instead of using a pseudoinverse approach, consensus-based partial coherence aims to improve the quality of the ill-conditioned cross-spectral density matrix $C_{\text{cond}}^{ij}$. It does this by using group averages for conditioning, instead of the $N - 2$ individual channels. The method is illustrated in Fig. 4.

As before, data (purple in Fig. 4) and reference data (green) were wavelet transformed and auto- and cross-spectra were computed (Fig. 4A). The auto- and cross-spectra then entered a permutation-based conditioning procedure. In each permutation, the $N - 2$ channels outside the pair under investigation were randomly assigned to $M$ groups (Fig. 4B). The auto- and cross-spectra were averaged within the groups, reducing the size of the cross spectral density matrix for every time-scale pixel from $N \times N$ to a much smaller $(M + 2) \times (M + 2)$ matrix. The pair's consensus was then conditioned using the reduced matrix. Because the reduced matrix was of high quality (i.e. high reciprocal condition number) it was inverted using the standard (LU decomposition based) inverse method.

The groupings used for matrix reduction were identical for dataset and reference and Z-scoring was performed for each permutation individually. This random-conditioning procedure was repeated 50 times. In order to find appropriate parameter settings we studied the quality of the consensus of 10 random pairs over a range of 5–100 permutations. Fig. 5B shows that the error in consensus decreased with more permutations, but this decrease flattened for over 30 permutations, suggesting that using 50 permutations results in reliable outcomes while limiting computation time. Other quality measures based on Z-score and qualitative outcome supported the data from Fig. 5B (see Fig. S4).

The number of groups, $M$, used for analysis of both model and sEEG data was set to 2–4 and was randomly chosen for each permutation. The consensus improved slightly with the number of groups (Fig. 5C, see also Fig. S4). However, a low number of groups allowed for the analysis of datasets with a few trials (see Fig. 1D), making it possible to track changes in network dynamics that occur within a few trials. The number of channels in each group was randomly chosen. The only restrictions imposed were: 1) groups could not be empty; and 2) permutations could not be identical, though the number of groups and the number of channels in each group were allowed to be the same across permutations.

Then, for each time-scale pixel the number of permutations with significant outcomes was counted: if more than 90% (i.e. 45 out of 50 permutations) was significant, we concluded that there was ‘consensus’ for partial coherence. The motivation for choosing 90% as a cut-off is shown in Fig. 6 (model) and Fig. 9 (human data) and is further motivated in the Results section.

The consensus captures the reliability with which a pixel is significant, but does not incorporate the amplitude of the effect. For example, a pixel can be significant in 95% of the permutations and hence reach consensus, independent of whether the coherence Z-score is 2.1, 6.0 or any other value. To estimate effect size, we averaged the Z-scores across permutations and ‘masked’ the average Z-score by the consensus. The mask weight was 1 for all values above the consensus threshold, and decreased linearly to zero for consensus values below the threshold (Fig. S1). We averaged the Z-scores across different permutations rather than the actual partial coherence values, to eliminate the potential influence of differences in mean partial coherence or coherence bias (Lachaux et al., 2002) between the different random groupings. For similar reasons, we compare only Z-scored values for the COH, cPCOH and PINV methods throughout this paper.

**Reference distribution for the human stereoEEG dataset**

To assess the significance of the acquired wavelet coherence and partial coherence, it was Z-scored against a reference distribution. In this paper, we aim to compare the networks that coherence and partial coherence methods produce to each other. To simplify the interpretation of this comparison, we tested the data from the two conditions of the task independently against the same, neutral baseline, instead of contrasting them to each other. We obtained the reference distribution by shuffling trials from different rules (Buschman et al., 2012): while maintaining the original data for the first channel in the pair, the data of the second channel was replaced by a random trial from the other rule (Fig. 4C). Partial coherence was then computed for this hybrid dataset. This procedure extracts those pairs that are differently modulated during the two conditions, independent of partial coherence.

![Fig. 5. Average error of cPOH-consensus for different numbers of repetitions (A), permutations (B) and number of channel groups (C). Error is computed against the largest number of repetitions, respectively permutations for A and B, and for the total dataset with all group sizes for C. Blue bars indicate the value that was used throughout the rest of the paper. Data from 10 randomly chosen pairs from patient P102.](image-url)
This rule shuffling procedure was repeated 100 times, hence obtaining 100 time- and scale-specific reference data points for each pair, as the quality of the outcome showed little improvement when using more than 50 repetitions (Fig. 5A and Fig. S4).

**Pixel-based assessment of performance**

For the model dataset, the outcome of the partial coherence calculation was compared to a ground truth dataset, which was identical to the test dataset, but did not have any shared inputs to the pair under study (Fig. 3C). This allows us to test the behavior of pseudo-inverse-based and consensus-based partial coherence under a range of conditions. We compared the qualitative (significant or not) and the quantitative (Z-score) output of both partial coherence methods to the coherence of the ground truth pair for every time-scale pixel.

The quantitative performance of the two partial coherence methods was computed as the mean squared error between the Z-scores of PCOH and the ground truth COH pixels. For this analysis we only used significant pixels, since consensus-based partial coherence Z-scores were only computed for pixels that reached consensus (see Section 2.4).

The qualitative performance of the methods was computed by counting the number of pixels that fell in each of four groups (see also Fig. 6):

1. significant ground truth coherence and significant partial coherence: the ‘kept’ or true positive (TP) pixels;
2. significant ground truth coherence, but no partial coherence: the ‘lost’ or false negative (FN) pixels;
3. no significant ground truth coherence, but significant partial coherence: the ‘new’ or false positive (FP) pixels;
4. no significant ground truth coherence, and no significant partial coherence: the true negative (TN) pixels.

With these counts, we computed two complementary measures, precision and recall, as is commonly done in both the medical and machine learning fields (Powers, 2011). The precision is the fraction of true positives over the total number of partial coherence positives. A high precision means that the partial coherence method filters out the shared inputs reliably. The recall, or the number of true positives over the total number of ground truth positives, indicates how many of the positive responses were identified by partial coherence. The higher the recall, the more complete the outcome is.

Precision and recall were also used to determine the impact of the consensus threshold on the consensus-based partial coherence. As precision and recall generally vary in opposite direction with threshold (Hernandez-Orallo et al., 2012), we computed their geometric mean, known as the G-score (Flach, 2003), to aid in the identification of the appropriate threshold.

For the sEEG dataset, no ground truth data was available. Therefore, the performance was assessed relative to the coherence computed for the same pairs. The impact of this choice on the interpretability of the quality measures is assessed in SI section 3.2 and discussed in the Results and Fig. 9.

**Blob-based assessment of performance**

The neural events underlying significant changes in (partial) coherence are expected to span several pixels in time and/or scale. We therefore assessed the performance of the partial coherence methods on the two-dimensional (time-scale) response patterns: ‘blobs’. Blobs were defined as areas of 11 or more connected pixels with a significant (partial) coherence (see SI section 2.4 on an assessment of the impact of this blob definition). Blobs were only used for further analysis if the center of mass of their pixels was within the region of interest (ROI), here defined as a frequency higher than 20 Hz and below 100 Hz, and timed in the period of 0–350 ms after cue or stimulus onset.

The process of blob extraction started with thresholding. For COH and PINV, all pixels with significant (α < 0.05, before FDR-correction) Z-scores were extracted, simplifying the time-frequency image to –1 for significantly negative Z-scores, 1 for significantly positive Z-scores and 0 for all other pixels. For cPCOH, thresholding was performed on the consensus ratio, with pixels that reached consensus set to 1 or -1 for positive or negative pixels, respectively, and to 0 for pixels that did not reach consensus. Pixels were grouped into blobs using a Connected-Component Labeling approach (Dillencourt et al., 1992). This 2-pass algorithm walks through the pixel space and compares all –1 and 1 pixels to the values of its four top and left neighbors (‘8-connectivity’). In the first pass, each pixel with value 1 or -1 receives either a unique label, if the pixel has no neighbors with the same value, or with the same label as its left neighbor. In the second pass, connected regions with different labels are merged, leaving unique regions with unique labels.

Subsequently, the peak Z-scores within each of the blobs were determined using a 2D peak finding algorithm. If more than one peak was found in a blob, the blob was split up by assigning each of the constituent pixels to the peak closest to it in a Euclidean sense.

For each partial coherence blob, overlapping blobs in coherence were identified, and vice versa, by multiplying the binarized time-scale images of COH with either cPCOH or PINV, and looking for the non-zero values. A blob was deemed ‘kept’ if at least ten of its pixels were present in both time-scale images, otherwise, the blob was considered ‘lost’ (see SI section 2.4 for how the results varied with the number of pixels used to determine overlap). COH blobs were ‘split’ if there was overlap with several cPCOH/PINV blobs, while a cPCOH/PINV blob with a field overlapping more than one COH blob was labeled as ‘merged’. Using the number of overlapping blobs we computed, as before, the precision, recall and their geometric mean, G-score.

**Correction for multiple comparisons**

The sEEG dataset contained several thousands of pairs of channels, each with hundreds of time-scale pixels and up to dozens of blobs. To avoid spurious results with this high number of comparisons, we entered all p-values of all pairs into a False Discovery Rate procedure, with q = 0.05 (Benjamini and Hochberg, 1995).

Throughout the paper, when pixel-based results are reported, the p-values corresponding to the Z-scores of all time-scale pixels were entered into the FDR procedure and only pixels that survived FDR-correction were used. When we report blob-based outcomes, the identified blobs (i.e. not the individual pixels) were corrected for multiple comparisons, using the following strategy. We assumed that the signals of the pixels in a blob resulted from a single underlying source. The peak Z-score of the blob was taken as the aggregate strength of this source, and compared to scores generated by random Gaussian fields (Chumbley and Friston, 2009). The p-value of this comparison was then used in the FDR procedure, at a q-value of 0.05. The p-values were computed as described in Friston et al. (1994), using functions included in the Statistic Parametric Mapping (SPM12) package.

The results presented in Figs. 7 and 9–11 in this paper were obtained after FDR correction. The results in Fig. 6 show data without FDR-correction. This figure shows the performance of the partial coherence methods in the model, and in order to avoid confusion over the cause of this performance, no additional analyses, such as FDR, were performed in obtaining the results.

**Statistics**

As discussed in Section 2.5, statistics for the coherence and partial coherence pixels were assessed using a Z-scoring procedure against a rule-shuffled reference dataset. Blob significance was assessed against a random Gaussian field assumption, as described in Section 2.8. Otherwise, significance of presented effects was assessed using permutation-based tests (Maris and Oostenveld, 2007). Many of the presented
distributions failed to meet the assumptions for normality, equal group size and/or variance, therefore requiring a permutation-based approach. Unless mentioned otherwise, a rank sum (U) was used as test-statistic and 5000 permutations were used to assess significance. An α of 0.05 was used for all tests.

**Blob clustering**

All blobs identified for the pairs in the frontal cortex of subject P102 were clustered based on the frequency and time coordinate of the center of mass, the frequency and time and sign of the peak. The frequency and time coordinates were scaled to the range of the region of interest, such that all five dimensions ranged between 0 and 1. The Euclidean distances between the blobs were entered into an Unweighted Pair Group Method with Arithmetic mean (UPGMA) hierarchical clustering algorithm (Day and Edelsbrunner, 1984). Clusters were defined by a maximum distance of 0.3 within clusters, so groups with distances larger than 0.3 were split into different clusters.

**Software**

All analyses were implemented in Matlab (R2014a, The MathWorks), using the Distributed Computing Toolbox (The MathWorks) for parallel computing. The stereoEEG data were loaded using the ‘edfRead’ function provided by The MathWorks (version 2.10, 2012, Brett Shoelson). The ‘spm_P_RF’ function and related functions from SPM12 (“SPM toolbox” n.d.) were used for the computation of blob peak values. The FDR implementation ‘fdr_bh’ by David Groppe was used (version 2.3, 2015) and peaks were identified using a peak finding implementation by Adi Natan (version 1.12, 2013). Otherwise, standard Matlab functions and custom code were used.

**Results**

**Performance of the two partial coherence methods for model data**

To assess the performance of cPCOH and compare it to the performance of PINV, we used both methods to analyze a model dataset. This dataset consisted of a pair of interest, which shared a direct connection, and received (shared) inputs from other channels in the network. We compared the outcomes of both methods to the ground truth dataset, in which the shared inputs were removed. We assessed their performance both on the level of individual time-frequency pixels, as well as for areas of connected pixels, or ‘blobs’. We also used the analyses to determine the appropriate consensus threshold for the cPCOH method.

All time-scale pixels for the model pair under study were checked for significant ground truth coherence as well as for consensus-based and pseudoinverse-based partial coherence. Each pixel therefore fell into one of four categories, as illustrated for cPCOH with the contingency table in Fig. 6A. Counting the total number of kept, new and lost pixels for each pair allowed us to compute the precision and recall of the methods (Fig. 6B). Precision signifies the reliability of positive (i.e. above consensus threshold for cPCOH or above Z-score threshold for PINV) outcomes of the methods, with a precision of 1 meaning that all positive pixels represent true interactions. Recall, conversely, signifies the number of positive pixels in the ground truth that are discovered by the cPCOH or PINV methods, and therefore shows how complete the picture painted by cPCOH or PINV is.

[Fig. 6. Comparison of the performance of cPCOH and PINV on model data. A: Contingency table for the qualitative scoring of pixels and blobs, comparing cPCOH (or PINV) with the ground truth model dataset. B. With the total number of kept, lost and new pixels for a pair, the precision, recall and their geometric mean, the G-score, were computed. C. The measures shown in B given for different consensus threshold of cPCOH (gray lines) and for PINV (orange line), against the numbers of trials in the dataset. D: Definition of activity ‘blobs’: an area of 11 or more connected pixels with a significant response. Blobs were said to ‘overlap’ between ground truth and cPCOH or PINV if they had 10 or more pixels in common. E. Similar to the pixels, the kept (i.e. overlapping), lost and new blobs were counted and precisions, recall and G-score computed. Color conventions as in C.]
Fig. 6C shows the precision and recall for the model dataset for both cPCOH and PINV, for different consensus thresholds of cPCOH and for different numbers of trials in the dataset. The precision of cPCOH increased with consensus threshold and outperformed PINV for all threshold values over 0.2. This was also the case for all the other choices of model parameters we evaluated, specifically connection strengths, number of connections, number of channels and signal-to-noise ratio (see Fig. S6).

Precision and recall cannot be optimized simultaneously: the more stringent a method (e.g. higher threshold), the higher the precision and the lower the recall will be. This was also the case for the model dataset. While precision increased with consensus threshold, recall went down, going below the PINV value for thresholds of 0.8 and above. To determine a potential optimum for both precision and recall, their geometric mean, known as the G-score, was used (Flach, 2003). The G-score increased with consensus threshold (Fig. 6C). However, all thresholds performed equally or better than the PINV method (the 0.9 threshold performed similar to PINV, p = 0.62, permutation test).

In addition, cPCOH outperformed PINV in estimating the effect size. Here, effect size was taken to be the Z-score of significant pixels. The mean squared Z-score error relative to the ground truth data was substantially lower for cPCOH for most model parameter values, even at high thresholds (Fig. S8). PINV only outperformed cPCOH for very strong connections and signal-to-noise ratios over 1.2.

While pixel-based performance gives a good indication of the accuracy of cPCOH, it assumes independence between the activation of timescale pixels, which is not valid for the model, and is invalid for the sEEG dataset as well. Instead, an activation usually spans several neighboring pixels, creating areas of activations or ‘blobs’ (Fig. 6D). We analyzed the blob-based performance in the same way as for the pixels: counting the number of kept, lost and new blobs in each pair compared to the ground truth dataset and computing precision, recall and G-score.

For blob performance there also was a trade-off between precision and recall: a higher consensus threshold increased the precision, hence improving the reliability of the blobs found by cPCOH, but reduced the recall, hence increasing the number of incorrectly removed blobs (Fig. 6E). However, unlike the pixel-based performance, cPCOH outperformed PINV for all thresholds and for all parameter ranges (see Fig. S7). For high thresholds, precision was around 0.9, meaning that 90% of cPCOH blobs were true interactions, and cPCOH blobs were twice as reliable as PINV blobs. The decrease in recall with increasing threshold was less pronounced for blobs than for pixels, causing the G-score to be approximately independent of consensus threshold. The blob-based G-score of cPCOH varied between 0.6 and 0.8 for different parameter values (see Fig. S7 for other model parameters), while the PINV never reached G-scores over 0.4, and hence the cPCOH method overall performed twice as well as the PINV.

Importantly, the performance of the cPCOH method was almost independent of the number of trials in the dataset. By contrast, the PINV produced an increasing number of spurious responses with a decreasing number of trials, leading to an artificially high pixel-recall and low precision around 100 trials. For less than 100 trials, PINV collapsed and produced no significant responses anymore. On the other hand, cPCOH precision remained stable, even when computed for only 5 trials, though the recall dropped when fewer than 50 trials were used. This relative insensitivity to trial number shows that the consensus-based partial coherence can be used for datasets with few repetitions, and that two datasets with different numbers of repetitions can be compared to each other.

The model results also give us insight into the applicability of the partial coherence methods for different networks and modalities. As expected, the partial coherence performed better for stronger connections between the pair of interest, while performance went down when shared inputs increased in strength or power (see Fig. S6, S7 and S9A). Performance remained high when the shared input had a similar frequency compared to the pair of interest (Fig. S9B). Higher similarity between channels, due to stronger connections in the rest of the network (parameter \( c_i \)), or a higher distribution of the shared input to the rest of the network (parameter \( N_{\text{shared}} \)), increased performance of cPCOH, but not PINV.

The blob-based G-score did not change with consensus threshold across parameter values (Fig. 6 and Fig. S7), which signifies that the cPCOH method provides the experimenter with the freedom to prioritize either precision or recall through choosing an appropriate threshold. When recall is prioritized, a low threshold should be selected, while for cases that favor a high precision, a high threshold is needed. As stated in the introduction, for the analysis of the data obtained with the sEEG method we aimed to remove as many of the shared inputs from the dense connection matrix obtained from coherence analysis as possible, i.e. we aimed for a high precision. We therefore used a high 0.9 threshold for the remainder of the paper (but see the Discussion section and Fig. S13 for the implications of this choice).

Response statistics of partial coherence in the model dataset

To allow for a comparison of model and human sEEG data, we entered all simulated pairs into an FDR correction procedure identical to that applied to the sEEG data. In Fig. 7 we compare four sets of blobs that survived FDR control. The blobs in the ground truth data (GT), those in the coherence of the test data (COH) and those filtered via the partial coherence, calculated using the consensus method (cPCOH) or the pseudoinverse (PINV).

In panel A1 the blobs in COH are shown labeled according to what the PCHO filtering does to them, with the left bar representing cPCOH filtering and right bar PINV. COH blobs could be lost, meaning they were not present after PCHO filtering (white). To assess whether this PCHO filtering was correct, i.e. representing shared inputs, the lost blobs were compared to the corresponding ground truth blobs. When lost COH blobs were present as ground truth blobs, they were wrongly filtered out and hence they constituted false negatives (FN, indicated in the figure by the hatched pattern). When the lost COH blobs were not present as ground truth blobs they were correctly filtered out and should be considered as true negatives (TN). On the other hand, COH blobs could be kept after the PCHO procedure, either correctly according to ground truth, constituting true positives (TP), or incorrectly, in which case they were false positives (FP). When comparing cPCOH to PINV, the hatched parts are seen to be slightly larger in the PINV case than in the cPCOH case, i.e. there were more FP and FN produced by PINV than by cPCOH, indicating that cPCOH outperformed PINV in this aspect (49% of blobs incorrect for cPCOH; 54% incorrect for PINV).

Panel A1 only accounts for the blobs that were already present in the COH set, not for new blobs that were uncovered by either of the PCHO procedures. In panel A2 we show the classification of the blobs that survived PCHO filtering, which allows the quantification of these new blobs, which are indicated in gray. Like in panel A1, the kept blobs, i.e. those that already existed in COH, are colored blue for cPCOH and orange for PINV. The hatched parts of the bars again represent the kept blobs that were incorrectly kept (33% for cPCOH; 49% for PINV) and the incorrectly added new blobs (72% for cPCOH; 90% for PINV), which together form the FP cases. Comparing PINV to cPCOH, we see that the hatched parts for PINV are substantially larger, showing that PINV produced many more false positives than cPCOH. Overall, 73% of PINV blobs were false positives, showing that PINV is not appropriate to discover blobs that represent true direct interactions.

Furthermore, the ground truth blob set contained blobs that were not present in COH, and that were, in addition, also not added as new blobs by PCHO. These blobs represent another source of false negatives and are shown in panel A3. For both PCHO methods there were substantial number of these FN, but there were substantially fewer of these missing blobs for cPCOH (16%) than for PINV (26%). Hence, in this aspect cPCOH also outperforms PINV.

These data show that cPCOH produced fewer false positives and
Fig. 7. Blob statistics for the model dataset. The dataset was analyzed in the same way as the human dataset: all simulated pairs were taken together, entered into FDR-correction and subsequently compared to the coherence of the same dataset. A: Numbers of kept (colored), lost (white), new (gray) and missing (black) blobs for coherence and partial coherence blobs obtained via either cPCOH (blue) or PINV (orange). The striped parts of each bar indicate the blobs that were incorrectly categorized by the PCOH methods. The black bars indicate blobs that were missing both in the COH and the PCOH data. B: Area (i.e. the fraction of time-frequency pixels) of GT blobs (left) and COH blobs (right) that was recovered by the PCOH methods. cPCOH blobs recover a larger part of the GT blobs than do PINV blobs. Dashed lines indicate medians.

fewer false negatives and therefore performed better than PINV. When partial coherence has to be computed on a noisy data set, the cPCOH procedure should be used.

When comparing panel A1 and panel A2, we note that the size of the colored bars in panel A2 are different than in panel A1, so the number of kept COH blobs is different than the number of kept PCOH blobs. An increase in panel A2 as expected that COH blobs are split in two or more blobs by the PCOH procedure, while a decrease happens when several COH blobs are merged. For cPCOH, the blue bar in panel A2 was smaller, indicating blobs were merged. On the other hand, the number of blobs after PINV (the orange bar) was nearly double that in panel A1, indicating that many COH blobs were split, creating many false positives. This large difference in the number of kept blobs, as well as number of new blobs, can be identified independent of a ground truth. We will therefore use these indicators in the next section to compare the performance of cPCOH and PINV for the human sEEG recordings, where a ground truth is not available.

In addition to identifying the correct blobs to keep, the PCOH procedures should also occur as much of the blob's time-frequency area as possible. PINV blobs that were correctly kept usually recovered only part of the blob area of the ground truth responses (Fig. 7B, median maintained blob area was 0.68), while cPCOH blobs maintained significantly more of the GT blob area (median 0.91; rank-based permutation test: p < 0.0002). Similar results were found when COH blob area was compared (Fig. 7B right panel, p < 0.0002).

Which COH blobs are filtered and which are kept? We found that COH blobs that were kept, were slightly bigger than lost blobs, while no preference for stronger blobs (i.e. higher peak Z-values) was found (Fig. S10A). On the other hand, cPCOH blobs were smaller than COH blobs, and both the size and height of the cPCOH blobs matched the GT (Fig. S10B). These results show that the cPCOH method correctly estimates blob size and maintains the peak Z-score of the correct blobs.

Assessment of the two partial coherence methods on human sEEG data

We also applied the coherence, consensus-based partial coherence and pseudoinverse-based partial coherence methods to a sEEG dataset comprised of 903 pairs of leads located in the frontal cortex of patient P102. These data were recorded for the purpose of studying how cortical networks change dynamically during task-switching (see Section 2.1). However, here we use these data solely for assessing the suitability of cPCOH for analyzing experimental data and to determine whether it can extract sparser network connectivity patterns that simplify interpretation and further analysis. The subsequent functional analysis and neuroscientific interpretation of these data will be reported in a future publication.

As for the model, in some sEEG pairs, the majority of the coherence blobs were filtered away by cPCOH, as in the example in Fig. 8A. Here, only one of the blobs survived (white solid box). In other pairs, most coherence blobs were either preserved by cPCOH, or some new blobs were introduced, as for example the blob at 40 Hz in Fig. 8B (white dashed box).

The outcome of the cPCOH method depended on the choice of consensus threshold. To test whether the performance on the human sEEG dataset varied in a similar way as for the model data, we computed the precision, recall and G-score for the frontal cortex data from P102, similar to the analysis shown in Fig. 6. Since the ground truth was not available for the human data, we used COH as a reference. This change in reference is likely to affect the values of these quality measures. We quantified the impact of this change of reference in the model, by computing the model data from Fig. 6 with COH as a reference (Fig. S11). We found that for the model, a COH reference slightly over-estimated precision, recall and G-score when computed on a pixel-basis, but not on a blob-basis, suggesting it gives a valid impression of the relative performance of the cPCOH method. However, note that when using a COH reference, the values that can be obtained for the precision and recall are now limited by the fraction of blobs that have to be filtered out. For example, when 40% of the COH blobs have to be filtered and the PCOH performs perfectly, the found recall referenced to COH would be 60%. We therefore have to be careful when interpreting absolute values of the recall. However, trends for recall, precision and G-score can reliably be used to compare cPCOH to PINV, and to compare the methods' performance in experimental data and the model.

The human sEEG showed the same trends as the model, with precision increasing with consensus threshold, whereas recall decreased with threshold (see Fig. 9D). However, the precision and recall did not vary linearly with threshold, as they did for the model, and the biggest changes occurred above a threshold of 0.8. Unlike the model, the blob-based precision (Fig. 9E) initially went down with consensus threshold, to improve again when exceeding a threshold of 0.8. This difference was due to the use of an FDR-correction, and computing the measures before the FDR-correction resulted in outcomes similar to the model (Fig. S12). As for the model, we prioritized high precision and chose a consensus threshold of 0.9 for further analyses. Consensus threshold had only a
small impact on the number of kept, lost and new blobs after FDR (see Fig. S13).

The key insight afforded by these analyses is that for blobs the overall quality, as represented by the G-score, is relatively insensitive to variations in the consensus threshold because the precision and recall vary mostly in opposite ways. Hence, the data-analyst can adjust precision or recall to match the specific demands of the analysis task at hand, which corresponded for the experimental dataset analyzed here to a precision range between 0.8 and 1.0.

How does consensus-based partial coherence impact the identification of functional connections in the human sEEG data? In the model, we identified two trends that allowed us to compare the performance of cPCOH and PINV without a ground truth: 1) the fewer new blobs were found the better the performance, as the majority of new blobs were false positives; 2) obtaining a large number of kept PCOH blobs compared to the number of kept COH blobs was a sign of a high false positive rate.

In the human sEEG dataset, we found that cPCOH filtered out the majority of COH blobs: about three-quarters of the coherence blobs were lost (white in Fig. 9A&B). About one-third of the remaining blobs were split into two or more cPCOH blobs (dark blue in Fig. 9A&B). Conversely, most of the cPCOH blobs overlapped with one or more COH blobs (blue in Fig. 9A right), and the number of kept cPCOH blobs was the same as the number of kept COH blobs. Only 20% of the cPCOH blobs did not overlap with a COH blob and were therefore new, though a fraction of these blobs overlapped with a COH blob that was removed by the FDR, and was therefore ‘rescued’ by cPCOH (dark gray in Fig. 9A, right). There were no substantial differences in these statistics between patients (compare Fig. 9A and B. However, the cPCOH method filtered out fewer blobs and produced more new blobs in a randomly chosen set of pairs comprised of channels located outside of the frontal cortex of Patient 102 (Fig. S15A), suggesting that the outcome of filtering by partial coherence analysis is non-uniformly distributed across the brain.

PINV filtered a similar number of COH blobs, but produced a substantially higher number of partial coherence blobs. As in the model, the number of kept PINV blobs was higher than the number of COH blobs. PINV also introduced a large number of new blobs. Without FDR-correction, over half of the PINV blobs were ‘new’ (Fig. S19A), which was also the case for the model, suggesting PINV performance on the experimental dataset was as poor as for the model data. As was shown for the model, kept PINV blobs overlapped significantly less with their corresponding COH blob(s) than cPCOH blobs (Fig. 9C, median was 0.22 for PINV, 0.42 for cPCOH; p = 0.0002), suggesting that PINV recovered fewer of the true positive pixels than cPCOH.

We found that the characteristics of kept, lost and new blobs differed in the sEEG dataset, as they did for the model data. Kept COH blobs had higher peak Z-scores and were bigger than the lost blobs (Fig. S15B&C), though lost COH blobs had a bigger peak height spread. cPCOH reduced the size of kept blobs, as was also seen in the model, and peak Z-scores were lower for kept cPCOH blobs than their COH counterparts.

Taken together and consistent with the insights obtained from the analysis of model data, these results show that cPCOH outperforms PINV in terms of a reduced number of false negatives and especially a reduced number of false positives. More importantly, because cPCOH removes a large part of the coherence blobs, it could potentially facilitate further analysis by prioritizing those blobs that are most likely reflecting direct interactions. Whether or not this really aids interpretation depends on
whether, from a functional perspective, the remaining blobs are more ‘useful’. The fact that these remaining blobs have systematically larger sizes inspires confidence that the answer is affirmative. In the following section we analyze whether those differences, for the remaining connections, are also reflected in more structured spatial patterns.

**Direct connections decrease with distance between recording sites**

We counted the number of lost and kept COH blobs and the number of new cPCOH blobs for each of the 903 pairs from the frontal cortex of Patient 102. The pairs were subsequently divided into three categories, based on whether their blobs were mostly kept, mostly lost, or mostly new (for definitions see Fig. 10A). When pairs were referenced according to the anatomical location of their constituent channels (Fig. 10B), clusters appeared, indicating that the behavior of pairs containing close-by channels was similar and hence that some areas shared more inputs than others, which were subsequently removed by the cPCOH procedure. These results are intuitive – direct connections are often present between nearby cortical areas, constituting a deterministic, non-random pattern, whereas long-range connections follow other structural non-random patterns.

The number of blobs per pair correlated significantly with their geodesic distance across the cortex (Fig. 10C), reflecting a systematic variation in connection strength, the number of frequencies used and/or the number of interactions during the trial interval. This was the case for COH, PINV and cPCOH alike. Overall, the number of blobs was higher for pairs of close-by channels, and lower for pairs that spanned a large part of the cortex. This is in agreement with previous work in monkeys, where the aggregate strength of direct connections have been shown to decrease with distance (Markov et al., 2013a,b). The correlation with geodesic distance was stronger for cPCOH than for COH, since lost blobs only weakly satisfied this relationship (Fig. 10D). This again indicates that the removal of blobs by the cPCOH procedure is not random, but reflects important biophysical realities.

Blobs were smaller for longer geodesic distances: both the frequency range of the blobs, their time range and the number of pixels in the blob decreased with distance. This relationship was more pronounced for cPCOH and PINV than for COH.

The correlation with distance was consistently found across rules (data for orientation rule is not shown here), brain areas (see Fig. S16) and patients (Fig. S18). The results are summarized in Table 1. In all cases, pairs with smaller geodesic distances had more kept blobs and correlation coefficients were substantially more negative for kept blobs than for lost and new blobs. These results suggest that direct connections...
Correlations across pairs between the numbers of kept, lost or new blobs and the geodesic distance between the members of the pair, for two datasets from Patient 102 and the frontal cortex of Patient 103. Non-significant correlations (alpha = 0.05) are not shown.

|               | COH | cPCOH |
|---------------|-----|-------|
|               | Lost| Kept  |
| Patient 102 – Frontal cortex | n.s. | 0.57  |
| Patient 102 – Rest of the brain | 0.24 | 0.58  |
| Patient 103 – Frontal cortex   | 0.29 | 0.56  |
|                             | Kept| New   |
| Patient 102 – Frontal cortex | 0.47 | 0.17  |
| Patient 102 – Rest of the brain | 0.52 | 0.23  |
| Patient 103 – Frontal cortex   | 0.48 | n.s.  |

are more local than spurious or shared connections and further that the cPCOH measure is effective in uncovering these more relevant connections.

From sEEG pairs to networks

To illustrate the impact of filtering using cPCOH on the identification of functional networks, we compared all identified blobs based on the sign of the peak (negative or positive Z-score), the location of the center of mass of all pixels and the location of the peak. Similar blobs were grouped into ‘clusters’ (see Methods section 2.10). This method allowed us to identify networks with a specific activation pattern in time and frequency. The clustering was performed for both COH, PINV and cPCOH blobs.

An example cluster, with similar time- and frequency characteristics for COH, PINV and cPCOH, are shown in Fig. 11A. The activation matrices in Fig. 11A reveal the pairs that were involved in the cluster. The activation matrices of the cPCOH and PINV clusters were much sparser than the COH activation matrices. The PINV matrix had minimal overlap with the COH blobs and most connections in the cluster were new (gray) and scattered throughout frontal cortex. On the other hand, cPCOH showed more similarities between neighboring pairs. The pairs showing mostly kept blobs (blue) and pairs showing mostly new blobs (gray) were grouped together. Regions involved in specific networks were therefore better defined, more spatially restricted, significant and with high accuracy designated as direct. As a further example, the COH cluster shown in Fig. 11B covered almost the entire frontal cortex of P102, while cPCOH identified a few ventral connections within lateral Prefrontal Cortex (lPFC) together with a group of anterior lateral PFC – medial PFC connections as the source of this activation. PINV failed to produce a cluster at the same time and frequency as COH and cPCOH, and the cluster with the highest time-frequency resemblance showed dense and scattered connections.

Discussion

In this paper, we set out to develop a robust partial coherence analysis pipeline for a stereoEEG (sEEG) dataset, typically comprised of more than one hundred recording channels in the cerebral hemispheres. For these data there is an embarrassment of riches – each pair of channels has many functional connections, so far making it impossible to extract meaningful information on the dynamics of cognitive processes. Methods used to identify functional connections are prone to errors, either through finding indirect connections or correlation due to shared inputs (i.e. false positives), or by missing blobs (false negatives). We have developed a new statistical tool that can make sense of these difficulties, but important data by filtering out the blobs representing indirect communication and reduce those that correspond to false positives. We proposed a consensus-based approach (cPCOH), in which the effective dataset size was reduced by averaging within random groups of channels. The influence of those group averages on a pair of interest was subsequently partialized out. By repeating these random groupings (permutations), a consensus outcome was established. We tested the performance of this consensus-based approach in a modeled dataset and compared it to the partial coherence method using a pseudoinverse (PINV). We used partial coherence analysis for the identification of response blobs and the network created by the pairs of channels that contribute to them in sEEG data from two human subjects. We have conducted an extensive validation procedure, which we discuss in the following, together with a comparison to alternative methods and application to other data modalities.

The interpretation of coherence and partial coherence

Coherence and by extension, partial coherence, is a measure of linear temporal correlation within a frequency band. Hence, when the brain regions sampled by two (stereo)EEG electrodes are connected by a direct connection that causes them to be correlated, coherence is expected to be high. However, this inference does not necessarily work the other way around: high (partial) coherence does not imply a direct neural connection. Instead, high coherence demonstrates a distributed effect, that can be caused by both direct and indirect neural connections, common inputs, or non-neural causes such as volume conduction and entrainment or other forms of locking to an external stimulus. Partial coherence removes common sources, including volume conducted signals, as long as this common information is represented in other channels of the dataset. In the realistic scenario that the recordings do not fully cover the brain, directness can therefore not be established with certainty, as intermittent nodes in the network might be missing from the dataset. The benefit of partial coherence over coherence instead lies in the ability to localize the correlation to specific pairs of recording sites: when partial coherence removes a response pixel or blob we can be sure, within the precision of the method, that this correlation was redundant and originating from a common source. Such a filtering makes it easier to detect and interpret changes in network configuration over time or between conditions.

It is likely that interactions between cortical circuits cannot be fully captured as a correlation within a frequency band, for example due to differences in preferred frequency between the circuits (Ray and Maunsell, 2010), differences in attentional demand (Bosman et al., 2012), or interactions between temporal or spatial scales (e.g. cross-frequency coupling). Even when circuits have matching frequencies on average, their frequencies can still vary on single trials, as frequencies have been shown to vary on short time scales (Atallah and Scanziani, 2009; Lowet et al., 2017). In those cases, coherence and partial coherence will not correctly estimate the interaction between the circuits and the impact of a common input. When a more detailed assessment of the interactions in a triplet of signals is needed, for example into the exact amount of overlapping and unique information in each of the signals, additional analyses, like the recently developed partial information decomposition (Wibral et al., 2017) may be more appropriate. Partial coherence can be used to assess when and where such detailed analyses are appropriate.

Even if (partial) coherence or the precise division and synergy of information can be established, it remains to be shown that there is a causal link between the observed coherence and synchronization or information transfer between the underlying neural circuits. Coherence can be boosted by strong phase locking (Lachaux et al., 1999), which itself might be more accurately estimated using other methods (Lowet et al., 2016), while coherence also increases through high amplitude coupling. Coherence does not reflect the specific phase difference between the circuits, which has previously been shown to influence information transfer (Beservo et al., 2015; ter Wal and Tiesinga, 2017; Womelsdorf et al., 2007). However, recent modeling work has shown that low coherence is accompanied by low information transfer between oscillators (ter Wal and Tiesinga, 2017). However, even short burst of coherence can underlie transmission of information (Palmigiano et al., 2017), indicating that time-resolved coherence is a powerful starting point for studying information routing.
cPCOH outperforms PINV in various aspects on model data and sEEG data

To assess the performance of cPCOH, we generated a model for which the ground truth was known, and wherein the correct blobs were identified using coherence. Based on model simulations, we show that cPCOH removes the blobs representing indirect connections and has lower rates of both false positives and false negatives than the only alternative robust partial coherence method for these data, the pseudo inverse-based partial coherence (PINV) (Fig. 6). This result generalized across most network parameter settings explored in the model simulations. For high consensus thresholds, the typical precision of the blobs found by cPCOH was between 0.8 and 0.9. In contrast, precision scores for PINV were typically around 0.4 and never exceeded 0.6. This was primarily due to a high amount of ‘noise’ introduced by the low quality inverse, which produced extensive spurious activation, especially in smaller datasets with fewer trials. cPCOH performance was largely independent of the number of trials in the dataset, and can therefore be reliably used for small datasets and for the comparisons of datasets of different sizes. The cPCOH method correctly estimated blob size and peak Z-score of the correct blobs and cPCOH recovered more of the blob area than PINV (Fig. 7). Therefore, overall, the blobs identified by cPCOH were more precise and better matched to the ground truth shape than PINV.

The model has several limitations as a way of estimating performance on experimental data. First, the model is linear. It is not clear whether the experimental signal can be represented, over its entire dynamical range, by a linear model. Second, the non-stationarity in the model comprised of a simple instantaneous step in connection strength, representing an onset of the cue, stimulus, or behavioural response. In the experiment, this onset is expected to cause changes in neural activity, as well as in the effective connection strength between recording sites, which can be smeared out across time and may be modulated by cortical state changes. Therefore, the model simulations can be expected to provide insight into the trends obtained in response to variations in analysis parameters, but should not be expected to predict the quantitative performance accurately.

We found that cPCOH on real data indeed followed the performance trends found for the model, confirming the enhanced statistical quality compared to the PINV method. cPCOH substantially reduced the number of blobs compared to coherence analysis for both patients (Fig. 9). Pairs that lost most of their blobs were grouped together in cortical space. Furthermore, pairs of which the constituent channels were closer together had more and bigger blobs (Fig. 10), a result that was consistent across brain areas and patients. Hence, the pilot experimental data set shows that the remaining blobs are meaningful, as they generate non-random structural patterns (Fig. 11) and, when referenced to their anatomical location, their occurrence and strength fell off with distance between the channels making up the pair, matching expectations based on anatomical studies (Markov et al., 2013a,b).

Fig. 10. Spatial aspects of cPCOH filtering. In A&B, the performance of individual pairs is shown. Blue: most blobs within the pair are kept, few are lost or new; black: most blobs of the pair are lost; gray: more new blobs than kept blobs (in parenthesis is the number obtained when blobs rescued from FDR are not considered). Each line is a pair. B: Connection matrix of frontal cortex (patient P102) indicating the categories of the pairs by color. IPFC: lateral prefrontal cortex, mPFC: medial prefrontal cortex; OFC: orbito-frontal cortex. The dashed boxes indicated two examples of a cluster of pairs comprised of nearby channels that lost most blobs after cPCOH filtering. Panels C&D show the relationship between blob characteristics and the geodesic distance between the channels in the pair. C: Correlation coefficient between geodesic distance and the number of blobs per pair, and the frequency spread, time spread, and number of pixels in the blob. D: Number of blobs per pair plotted against geodesic distance of the pair, for lost, kept and new COH and cPCOH blobs.
Attention (Bosman et al., 2012), sensory input (Ray and Maunsell, 2010; Roberts et al., 2013) or cue and stimulus onset (Churchland et al., 2010). On the other hand, frequency bands found in spectra and coherence have been shown to change with task demand, such as stimulus onset, and at a higher frequency than the cluster in panel B. To ensure that the time-frequency blobs are an accurate reflection of functional networks in sEEG data, the type I error should not exceed 5% of all significant outcomes. In this paper there are two types of false positives and false negatives. First, (partial) coherence is a statistical quantity, meaning its estimation has uncertainty associated with it: a true coherence can be judged as non-significant (type II error or false negative) or a non-existing coherence can be judged to be significant (type I error or false positive) (Bokil et al., 2007). We controlled this error through FDR-correction (Benjamini and Hochberg, 1995). As we estimated the coherence in a time and frequency resolved way using wavelets at relatively high resolution, coherence values of neighboring pixels might be correlated. We therefore had to perform FDR correction on the blob level. As mentioned before, the results reported here have been obtained after controlling the FDR at 5%, meaning that the type I error should not exceed 5% of all significant outcomes.

The second source of FNs and FPs in this paper are mistakes made by the PCOH methods, for example, an indirect connection being marked as direct by PCOH would constitute a false positive. Here, we do not have the benefit of the FDR-correction and we cannot control for these mistakes in a statistical sense, hence we can only empirically estimate performance in terms of precision and recall according to the model-validated coherence-based reference method.

To ensure that the time-frequency blobs are an accurate reflection of the underlying network activation patterns while maintaining a reasonable computation time, the time- and frequency resolution of the analysis needs to match the resolutions of the activation pattern. To put this in a different context, imagine sampling an Alpine landscape: neither 1 mm nor 100 km would be an appropriate grid size. Since wavelet based coherence and partial coherence are not influenced, in a statistical sense, by the time and frequency resolution, one can freely choose the resolution using other factors: the type of data, research question, task design, and frequency band of interest. Here, we validated the choice of resolution by counting the number of small COH blobs (Fig. S5) and found very few blobs smaller than 10 pixels, indicating that the resolution was appropriate to capture the correlation landscape.

**Blob-based identification of functional networks in sEEG data**

Existing literature on monkey data suggests that different functional connections can use different frequency bands for communication (Bastos et al., 2014; Quax et al., 2017; ter Wal and Tiesinga, 2017; van Kerkoerle et al., 2014; Womelsdorf et al., 2014). This multiplexing in frequency can aid the distinguishability of such connections, which can otherwise be difficult to disentangle due to the large number and high diversity of neural populations that give rise to sEEG signals (Buzsáki et al., 2012). On the other hand, frequency bands found in spectra and coherence have been shown to change with task demand, such as attention (Bosman et al., 2012), sensory input (Ray and Maunsell, 2010; Roberts et al., 2013) or cue and stimulus onset (Churchland et al., 2010). Analysis therefore has to be conducted on short time windows that are shifted across the interval of interest to give a full account of the dynamics. Using the wavelet coherence rather than the ‘Fourier’ coherence as a basis for the partial coherence can effectively deal with this non-stationarity. The non-stationary character and the use of many frequency bands indicate that the atom of communication is the time-frequency blob. We therefore focused the validation of the cPCOH procedure on these blobs, rather than individual frequencies or times-of-interest and aimed to control the number of false positive and false negative blobs produced by the method.

To ensure that the time-frequency blobs are an accurate reflection of the underlying network activation patterns while maintaining a reasonable computation time, the time- and frequency resolution of the analysis needs to match the resolutions of the activation pattern. To put this in a different context, imagine sampling an Alpine landscape: neither 1 mm nor 100 km would be an appropriate grid size. Since wavelet based coherence and partial coherence are not influenced, in a statistical sense, by the time and frequency resolution, one can freely choose the resolution using other factors: the type of data, research question, task design, and frequency band of interest. Here, we validated the choice of resolution by counting the number of small COH blobs (Fig. S5) and found very few blobs smaller than 10 pixels, indicating that the resolution was appropriate to capture the correlation landscape.

**Optimisation of cPCOH for sEEG and other data types**

In this paper there are two types of false positives and false negatives. First, (partial) coherence is a statistical quantity, meaning its estimation has uncertainty associated with it: A true coherence can be judged as non-significant (type II error or false negative) or a non-existing coherence can be judged to be significant (type I error or false positive) (Bokil et al., 2007). We controlled this error through FDR-correction (Benjamini and Hochberg, 1995). As we estimated the coherence in a time and frequency resolved way using wavelets at relatively high resolution, coherence values of neighboring pixels might be correlated. We therefore had to perform FDR correction on the blob level. As mentioned before, the results reported here have been obtained after controlling the FDR at 5%, meaning that the type I error should not exceed 5% of all significant outcomes.

The second source of FNs and FPs in this paper are mistakes made by the PCOH methods, for example, an indirect connection being marked as direct by PCOH would constitute a false positive. Here, we do not have the benefit of the FDR-correction and we cannot control for these mistakes in a statistical sense, hence we can only empirically estimate performance in terms of precision and recall according to the model-validated coherence-based reference method.

How can we adjust the method such that for specific experimental data we can reach optimal G-scores? cPCOH has one main parameter to be set by the experimenter: the consensus threshold. This threshold presents a classical trade-off: how many false positives do we accept in order to minimize false negatives, and vice versa? For blobs, we showed that an optimum value between the precision and recall does not exist. The trade-off between false positives and false negatives therefore is a parameter that can be freely chosen.

We reported the G-score as a summary of overall performance. The model showed that performing analysis at the blob level (maximal G-score close to 0.7) yields more reliable results than at the pixel level (maximal G-score close to 0.5) (Fig. 6). The blob-based precision could be titrated from 0.8 up to 1.0 at the cost of moderate decreases in the degree of recall. If positive outcomes have to be highly reliable, as was our requirement here, the threshold should be increased, while if the positive
outcomes have to be complete, the threshold should be set to low values.

The experimental data fared less well on the overall G-score (Fig. 9), with maximal values close to 0.55 for blob-based analysis and 0.25 for pixel-based analysis, the primary reason being very low recall levels. The most likely explanation for this low recall is a systematic underestimation of recall due to using the coherence as a reference. When coherence is dominated by indirect connections, the majority of the COH responses have to be filtered out, limiting the recall to low values, especially when the performance of the partial coherence method is high. This is consistent with the observation that the recovered fraction of the COH blobs, that is, those pixels that are also present in the corresponding cPCOH, is lower than in the model (Fig. 9C). In addition, precision and recall followed the same trends with consensus threshold for both model and sEEG data set, with recall decreasing with consensus threshold, while precision was uniformly high. These trends suggest that the error rate of cPCOH was as low for the sEEG data as for the model.

In addition to consensus threshold, cPCOH performance depends on the number of repetitions, permutations and group sizes. Our analysis on the pilot experimental data suggests that these can be derived using internal controls, as the change in performance decreased while the parameter values increased (Section 2.5). The parameter values at which this convergence of performance takes place may vary between datasets and may be taken as a proxy for the quality/complexity of the data.

Can performance of the cPCOH method be further improved by separating the false positives from true positives and the false negatives from true negatives? Such a post-hoc selection would require a substantial difference in blob characteristics between TP and FP or TN and FN so they may be identified for individual blobs. For example, the model data suggest that false positive kept blobs are substantially smaller than their COH counterparts, while true positive kept blobs differ in size from the COH blobs only to a small extent (Fig. 7). If more characteristics exist, they together have the potential to assign a confidence score to each of the identified blobs and improve the outcome of cPCOH by identifying unlikely positives. In the Supplemental Discussion we provide suggestions for how such a confidence score could be formalized, however developing a filtering approach based on such a score would require additional model studies, which are beyond the scope of the current study.

**Applicability of the consensus partial coherence to other data modalities**

Methods to partial out shared inputs have proven their use in the analysis of MEG and EEG data, where leakage between channels or sources is a common problem. Leakage is usually dealt with by removing all zero-lag correlations, either by performing pair-wise partialization (Brookes et al., 2012) or more recently, multivariate zero-phase partialization (Colclough et al., 2015). Partial coherence moves beyond that and has a different aim: to also remove results due to shared inputs with an additional model studies, which are beyond the scope of the current study.

Applicability of the consensus partial coherence to other data modalities

Methods to partial out shared inputs have proven their use in the analysis of MEG and EEG data, where leakage between channels or sources is a common problem. Leakage is usually dealt with by removing all zero-lag correlations, either by performing pair-wise partialization (Brookes et al., 2012) or more recently, multivariate zero-phase partialization (Colclough et al., 2015). Partial coherence moves beyond that and has a different aim: to also remove results due to shared inputs with an additional model studies, which are beyond the scope of the current study.

In summary, the consensus-based partial coherence method provides a powerful filtering method for sEEG data with a large number of simultaneously measured channels. Compared to the coherence and the pseudoinverse-based partial coherence, its outcomes contain less shared inputs and less noise, respectively, highlighting a key advantage of the cPCOH method, that would thereby provide a more localized and precise representation of underlying neural responses. We therefore intend to use this method to identify switches in cortical network configuration following rule switches. The current results not only support that this method is applicable for this type of research question as well as for other analyses of sEEG data, but that it is also more broadly applicable to other modalities, including datasets with many simultaneous recording site and/or few trials.

**Declarations of interest**

None.

**Code availability**

The code used for the computation of the consensus-based partial coherence was written for Matlab (The MathWorks) and uses the Field-Trip data formatting (Oostenveld et al., 2011). The code is made available on https://github.com/marijeterwal/cPCOH.

**Acknowledgements**

This work is supported by funding from the European Union’s Seventh Framework Programme (FP7/2007–2013) under grant agreement nr. 600925 (NeuroSeeker), ERC Advanced grant nr. 323606 (ParietalAction), and European Union’s Horizon 2020 Research and Innovation Programme under Grant Agreement No. 720270 (HBP SGA1).

**Appendix A. Supplementary data**

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.06.011.

**References**

Albo, Z., Di Pricco, G.V., Chen, Y., Rangarajan, G., Truccolo, W., Feng, J., Vertes, R.P., Ding, M., 2004. Is partial coherence a viable technique for identifying generators of neural oscillations? Biol. Cybern. 90, 318–326. https://doi.org/10.1007/s00422-004-0475-5.

Atalha, B.V., Scanziani, M., 2009. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. Neuron 62, 566–577. https://doi.org/10.1016/j.neuron.2009.04.027.

Avanzini, P., Abbodlahi, R.O., Sartori, I., Caruana, F., Pelliccia, V., Casaceli, G., Mai, R., Lo Russo, G., Rizzolatti, G., Orban, G.A., 2016. Four-dimensional maps of the human somatosensory system. Proc. Natl. Acad. Sci. Unit. States Am. 113, E1936–E1943. https://doi.org/10.1073/pnas.1601889113.

Bastos, A.M., Schoffelen, J.-M., 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. 9, 1–23. https://doi.org/10.3389/fnsys.2015.00175.

Bastos, A.M., Vezoli, J., Bosman, C.A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J.R., De Weerd, P., Kennedy, H., Fries, P., 2014. Visual areas exert feedforward and feedback influences through distinct frequency channels. Neurosci. 85, 1–12. https://doi.org/10.1016/j.neuroscience.2014.12.018.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B. https://doi.org/10.2307/2346101.

Besserve, M., Loeve, S.C., Loeothesis, N.K., Schloophof, K., Panzeri, S., 2015. Shifts of gamma phase across primary visual cortical sites reflect dynamic stimulus-modulated information transfer. PLoS Biol. 13, 1–29. https://doi.org/10.1371/journal.pbio.1002257.

Bokil, H., Purpara, K., Schoffelen, J.M., Thomson, D., Mitra, P., 2007. Comparing spectra and coherences for groups of unequal size. J. Neurosci. Meth. 159, 337–345. https://doi.org/10.1016/j.jneumeth.2006.07.011.

Bosman, C.A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, A.M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., Fries, P., 2012. Attentional stimulus selection...
Wibral, M., Priesemann, V., Kay, J.W., Lizier, J.T., Phillips, W.A., 2017. Partial information decomposition as a unified approach to the specification of neural goal functions. Brain Cognit. 112, 25–38. https://doi.org/10.1016/j.bandc.2015.09.004.

Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A.K., Fries, P., 2007. Modulation of neuronal interactions through neuronal synchronization. Science 316, 1609–1612. https://doi.org/10.1126/science.1139597.

Womelsdorf, T., Valiante, T. a, Sahin, N.T., Miller, K.J., Tiesinga, P., 2014. Dynamic circuit motifs underlying rhythmic gain control, gating and integration. Nat. Neurosci. 17, 1031–1039. https://doi.org/10.1038/nn.3764.