Nonlinear brain connectivity from neurons to networks: quantification, sources and localization

Giulio Tani Raffaelli[a], **Stanislav Jiříček**[a,b,c], **and Jaroslav Hlinka**[a,b,1]

This manuscript was compiled on November 17, 2024

Since the first studies in functional connectivity, Pearson's correlation has been the primary tool to determine relatedness between the activity of different brain locations. Over the years, concern over the information neglected by correlation pushed toward using different measures accounting for non-linearity. However, some studies suggest that, at the typical observation scale, a linear description of the brain captures a vast majority of the information. Therefore, we measured the fraction of information that would be lost using a linear description and which regions would be affected the most. We considered fMRI, EEG, iEEG, and single unit spikes to assess how the observation scale impacts the amount of non-linearity. We observe that the information loss is reduced for modalities with large temporal or spatial averaging (fMRI and EEG) and gains relevance on more fine descriptions of the activity (iEEG and single unit spikes). We conclude that for most human applications, Pearson's correlation coefficient adequately describes pairwise interactions in time series from current recording techniques.

Functional Connectivity | nonlinearity | Mutual Information | fMRI | Electrophysiology

Significance Statement

In neuroimaging, as in other complex systems fields, the increasing interest in network inference by statistical dependencies (i.e. functional connectivity) invites advanced ways to quantify it. Various nonlinear measures, ultimately the Mutual Information, are used as alternatives to conventional linear Pearson's correlation coefficient. We systematically assess the amount and reliability of detectable nonlinearity of brain functional connectivity across imaging modalities and spatial scales. We demonstrate more pronounced nonlinearity in micro-scale recordings, while rather limited and unreliable in more accessible, noninvasive, large-scale modalities: functional magnetic resonance imaging and scalp electrophysiology. This fundamentally supports the use of robust and easily interpretable linear tools in large-scale neuroimaging, and important insights concerning the microscale connectivity nonlinearity, including the link to brain state dynamics.

Many complex real-world systems, such as the human brain, Earth's climate, or social or financial networks, do not easily allow probing their full intrinsic repertoire. Fully controlled laboratory experiments are unattainable, and the researchers increasingly complement traditional theoretical or experimental approaches with advanced data analysis on observational recordings of their natural behaviour. A common starting step in such endeavour is to capture the structure of statistical interactions between their constituent units. These interactions are then further analysed, e.g., through graph theory, machine learning, or statistical approaches.

In the field of neuroscience, the term functional connectivity is used for such statistical dependence between remote neurophysiological events (1, 2). It has been the key cornerstone of the paradigm shift stressing the importance of understanding the mechanisms of spontaneous brain activity dynamics (3, 4). Over the decades, a plethora of functional connectivity indices has been proposed and applied. These indices form a densely populated zoo of methods, differentially sensitive to various forms of the concerned statistical dependence. This variety ranges from the classical Pearson's correlation coefficient, assessing the strength of linear dependence between observed subsystems, to the principled use of mutual information. This is an entropy-based measure sensitive to any form of dependence, making it highly attractive for studying the interaction structures of highly non-linear or even chaotic systems. Over the years, a number of papers have reported the effectiveness of MI on different modalities such as EEG (5) or fMRI (6) and significant changes in MI that correlate with disease (7, 8).

Mutual information as a measure of dependence may thus appear as a clear method of choice due to its lack of theoretical assumptions. However, in practice, there are trade-offs that need to be taken into account. The issue of estimate accuracy, the need for longer time series (9), as well as computational demands, may favour more basic methods such as the Pearson's linear correlation coefficient. The pragmatic question the researchers (should) face is thus: Is the dependence pattern under study so exotic and non-linear as to warrant using more general methods (or even, in principle, mutual information) at the expense of decreased statistical power, increased computational demands, and other costs of more advanced dependence quantification methods?

While the presence of non-linearity in brain signal dependence structure is theoretically well established and beyond dispute, much less is known about its strength. A principled approach for its quantification in terms of extra-Gaussian information has so far only been applied to a single modality (10). The investigation was limited to functional magnetic resonance imaging (fMRI), and at a single

Author affiliations: [a]Institute of Computer Science, Czech Academy of Science, Prague, Czech Republic; [b]National Institute of Mental Health, Klecany, Czech Republic; [c]Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

GTR and JH designed the methodology and handled the validation and formal analysis. JH conceptualised and supervised the study, secured the funding, cared for the administration and managed resources. GTR wrote the software for the analysis and produced the visualisation and the first version of the draft. All authors conducted the investigation, curated the data, and reviewed and edited the draft.

No competing interests.

[1]To whom correspondence should be addressed. E-mail: hlinka@cs.cas.cz

PREPRINT

spatial signal averaging resolution given by the common, yet relatively coarse and arguably suboptimal Automatic Anatomical Labelling atlas (11). The results suggested that for this modality and at this spatial resolution, non-linear dependence contribution was likely negligible, supporting the practice of using Pearson's linear correlation for analysis of this type of coarse fMRI data. However, this still left the question of the relevance of non-linear dependences unanswered for a dominant proportion of neuroscientists, as the results could be only speculatively generalized to other neuroimaging modalities and spatial scales.

To at least partially fill this gap, we here apply the principled methodology for non-linearity quantification to neuroimaging data ranging across modalities from fMRI through scalp electroencephalography, intracranial electroencephalography, to single neuron (spike) recordings, adding analysis across orders of magnitudes of spatial averaging, and additional study of the sources, reliability, and localization of the observable non-linearity. Moreover, we provide an implementation in open code, allowing to quantify non-linearity in other datasets in neuroscience and beyond, which should provide a powerful tool also to researchers in other disciplines, following early successes of similar methodology in climatology (12) and financial networks (13).

Results

Probably the simplest, and in some sense canonical, form of dependence between quantitative variables is that of purely linear relation between two variables in a jointly normal (Gaussian) distribution, the strength of which is captured by the well-known Pearson's (linear) correlation coefficient. Indeed, linear correlation is sufficient to determine mutual information in a bivariate Gaussian distribution through the simple equation

$$I_{Gauss} = -\frac{1}{2}\log\left(1 - r^2\right) \ . \qquad [1]$$

It is exactly the deviation from the jointly Gaussian distribution that makes the use of Pearson's linear correlation coefficient potentially suboptimal, calling for the use of more advanced methods such as the universally sensitive mutual information. The strength of the deviation from Gaussian dependence pattern is thus the object of our interest. Moreover, we are not concerned with non-Gaussianity of the marginal distributions individually (as these can be easily treated by the use of monotonic nonlinear rescaling such as log-transform, or application Spearman's correlation instead of Pearson's), but particularly with such deviations from Gaussianity which concern the very relation between the variables, called copula, which entails the full characterization of the dependence structure, invariant with respect to any such bijective monotonic rescaling of marginals.

Note that to avoid too technical language, we shall use the terms "linearity' and "Gaussianity" (of distribution/dependence/interaction/pattern/process) interchangeably throughout this manuscript, although indeed in a strict sense the linearity is a bit wider term in particular contexts (one can e.g. imagine linear functional dependence as the best fit between two variables with non-Gaussian distributions or errors, or linearly coupled process with nonlinear driving noise).

We shall leverage that known statistical physics results that the bivariate Gaussian has the maximum entropy, and thus the minimal information (under Gaussian marginals), among the distributions with the same correlation, and thus every joint distribution that doesn't have a Gaussian copula (we shall call these distributions "non-linear"), has higher MI than expected from correlation by the formula Eq. (1). This allows us to define the distribution non-Gaussianity by $I_{extra}(X,Y) = I(X,Y) - I_{Gauss}(r(X,Y))$, or its normalized variants.

We identify two primary sources of non-linearity: intrinsic non-linearity and non-stationarities. Intrinsic non-linearity happens when the recorded samples are genuinely identically distributed, but they do not have a Gaussian joint probability distribution. For example, this can include a relationship between absolute values or out-of-phase synchronization phenomena.

We refer to non-stationarity when the samples are not identically distributed, and the source distribution depends on time. The samples will, in general, not be distributed according to a multivariate Gaussian, even if the source distributions are all Gaussians. Most estimators require the samples to be i.i.d. or at least from a stationary distribution. The non-stationarity, even of resting-state (rs) data, is well known and studied on its own as a potential source of insight into the brain functioning (14, 15). The effect's magnitude depends on the estimator and the non-stationarity's properties. Examples of non-stationarities are the switching between states—each associated with a different correlation, isolated bursts of activity, and a continuous drift of mean, variance, or correlation.

Along with these neural sources for the observed non-linearity, others can reside in the acquisition and pre-processing of the signal. For instance, non-monotonous transformations may easily result in an increase in observed non-linearity, and in particular, observation of nonlinearity between originally linearly related processes (it is illustrative to imagine the effect of taking, as a domain-specific preprocessing step, absolute value or square of Gaussian signals before probing their functional connectivity; similar but less straightforward effect is obtained by working on, e.g., Hilbert-transformed data or (band-limited) signal power time series, transformation much more common in neuroscience).

Strength of non-linearity. We begin by measuring the fraction of information in all connections not explained by a bivariate Gaussian copula. We call this measure Relative Non-Linearity (RNL).

We start from rs-fMRI and look at RNL over a range of region numbers and sizes to probe the effect of different degrees of spatial averaging. We observe a consistent presence of non-linearity for the different region sizes from the Craddock atlas (16). The fraction of MI not explained by the correlation (Fig. 1A) sits around 4% for all region sizes except for very few and large regions. At the same time, the non-linearity observed in the shadow (phase randomised) dataset remains below 2%, providing a measure of the bias in the absence of non-linearity. The difference in MI between empirical and shadow datasets is always significant, except for the atlas with ten regions (when applying strict Bonferroni correction for multiple comparisons).

PREPRINT

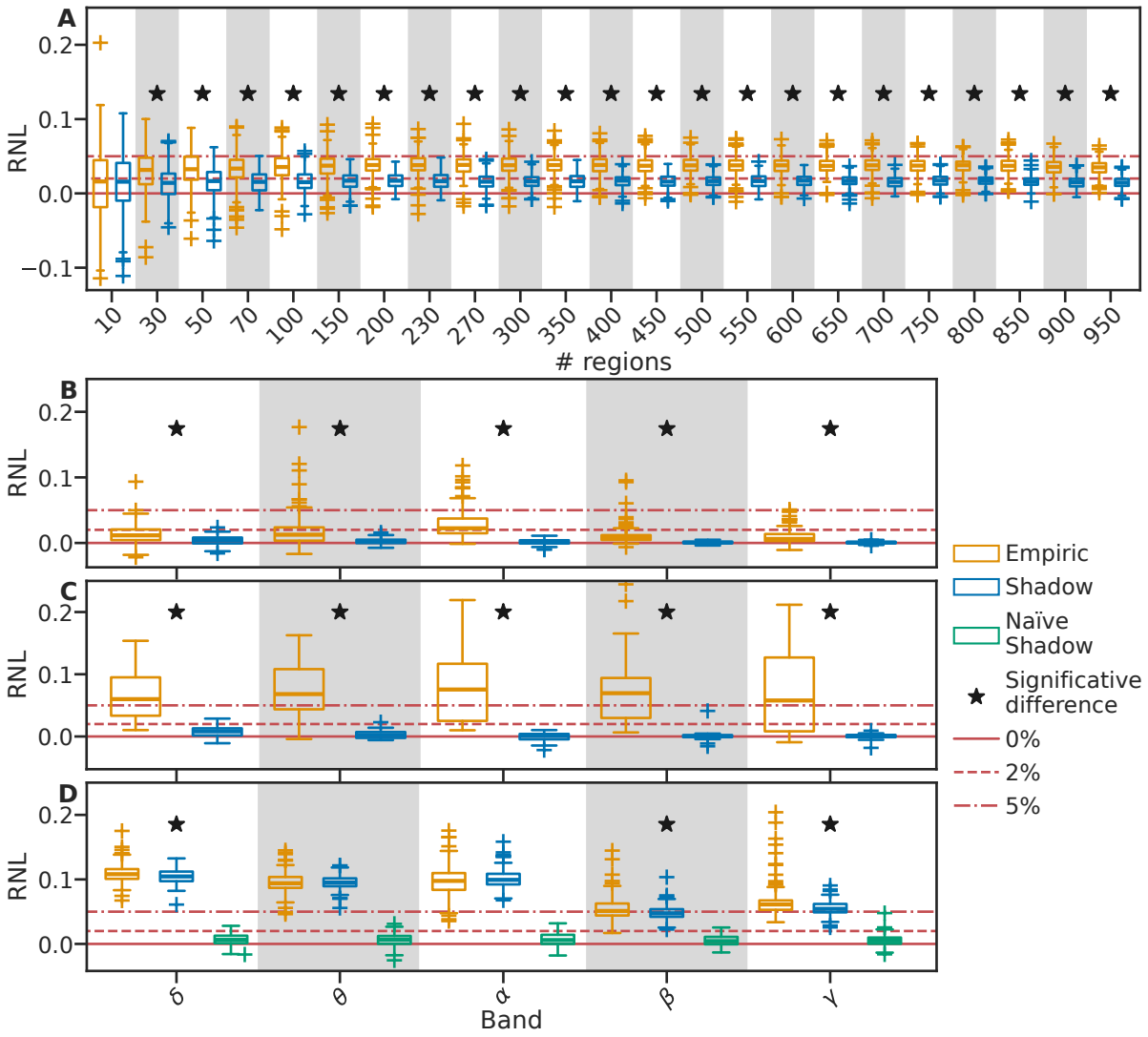

Fig. 1. Distribution over subjects of the Relative amount of Non-Linearity (RNL). A) fMRI data using Craddock parcellation and a varying number of regions, B) EEG scalp voltage, C) iEEG voltage, D) EEG band-limited power. The shadow dataset is a linear surrogate of the empirical one, see text for detail.

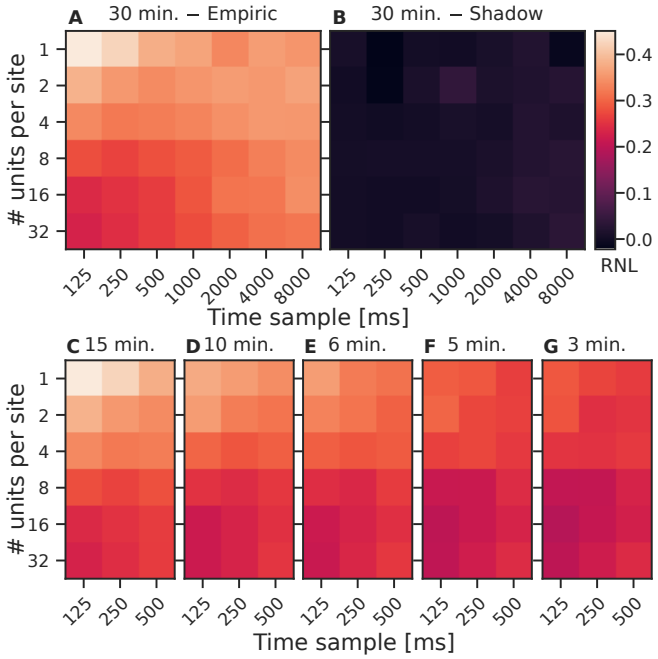

Fig. 2. Average RNL across mice varying the time average window size and the number of units. A-B) RNL computed over the full duration of the resting state experimental block, C-G) average of the empirical RNL computed over epochs of decreasing length.

In EEG time series, we observe a significantly higher presence of non-linearity (one-sided t-test, Bonferroni corrected) than in the shadow dataset for all frequency bands. In the α band, the amount is similar to what was observed in fMRI (Fig. 1B), while in the others, it tends to be even lower, in particular under 2%. The different number of samples available in each band explains the dependence between the frequency band and RNL in the Shadow dataset. See the Methods section and the SI for details.

In iEEG, non-linearity is more prominent than in the previous two datasets. For all bands, the empirical RNL tends to be above 5%. At the same time, in the shadow dataset, it is confined below 1% (Fig. 1C). Part of the non-linearity is explained by non-stationarities due to interictal epileptic activity (see SI). Moreover, the iEEG was acquired during natural activity (neither standard resting state nor sustained task), which might affect the stationarity of the sequences and their RNL.

Lastly, we looked at the RNL in single-unit spike rates in mice. In this case, we report the RNL averaged across sessions. The relative contribution of non-linearity is above 23% (Fig. 2A) in all cases, over the full range of temporal (between 125 ms and 8 s) and spatial/group size (between 1 and 32 units) averaging. We observe the highest values of RNL for the smallest group sizes and faster time scales. Conversely, averaging over larger groups always reduces the RNL, and averaging over more than 2 s has little effect with the RNL between 32% and 37% depending on group size. We found no correlation between the RNL and the number of regions in a given mice (ranging between 9 and 16). Thus, we present the results as the average across all mice. At the same time, the RNL never surpasses 5% in the shadow dataset (Fig. 2B). The magnitude and amplitude of the fluctuations increases for shorter time series as observed for lower frequency bands in EEG and iEEG.

Localisation. The amount of non-linearity is reduced for the more accessible modalities. In fMRI and EEG, high temporal or spatial averaging masks most of the non-linearity. However, it might still be relevant if localised in specific regions. The authors in (17) suggest localisation of non-linearity in the occipital region. We evaluated non-linearity's localisation, looking at regions that participate in consistently non-linear connections across subjects.

For fMRI data, we used the AAL90 atlas as it provides explicit anatomical labelling for the regions where the non-linearity might be localised. This dataset has a significant correlation ($p = .007$) between the localisation of non-linearity in the empirical and shadow datasets (see SI). Furthermore, correlation and localisation increase with reduced denoising steps in preprocessing. This suggests that most region-specific non-linearity is due to artefacts and is removed during preprocessing.

EEG data (Fig. 3) offer a different picture. Here, non-linear relationships are quite limited in the band θ (only 38% channel pairs) while present in more than 95% of channel pairs in other bands. In the shadow dataset, less than 0.8% of the relationships are consistently non-linear across subjects. At the same time, each region participates in connections with a non-random amount of non-linearity. For the slower bands δ, θ, and α, we observe that the degree of the occipital regions is higher (and in frontal regions lower) than expected from a random graph, suggesting a localisation of non-linearity. Conversely, for the high-frequency bands β and γ, the results show a predominance of non-linearity in frontal and temporal electrodes. This anterior-posterior gradient may be partially attributed to the spatial distribution of most prevalent sources in normal awake EEG (18), however more research is warranted.

Reliability. However small, accounting for non-linearity may still be beneficial if the additional information is stable over repeated measures. As the last test, we looked into non-linearity's reliability across sessions. We compared the subject ranks for each connection based on Total Mutual Information (TMI, i.e., accounting for non-linearity) and those derived from correlation.

As shown in Figure 1A, the presence of non-linearity in fMRI data is significant for all atlases with enough regions. However, the amount is limited and unreliable under our measure. The reliability for the Pearson's correlation is, on average, more than twice that of TMI (0.277 against 0.134). Moreover, if we estimate MI from correlation, losing the sign of the relationship, we drop to similar levels of correlation across sessions (0.153). In this case, however, Pearson's correlation predicts TMI in a second session even slightly better (0.008 or 5% increase in correlation) than TMI predicts itself across sessions, showing the practical advantage of using linear functional connectivity measure at this scale.

Running the same analysis on EEG and iEEG data offers a similar picture (see SI), which again excludes a clear advantage of TMI.

PREPRINT

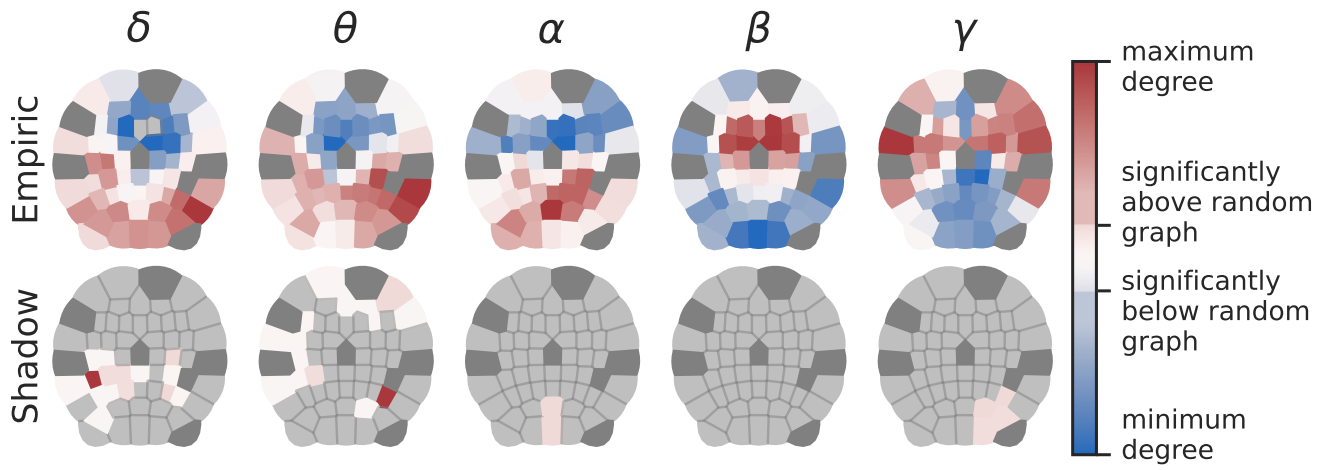

Fig. 3. Degree of the regions in a network weighted by the z-score of TMI compared to surrogates. Light grey regions have degree zero. Dark grey regions have been excluded from the analysis as often missing or corrupted. The representation is a Voronoi tessellation of the stereographic projection on the xy-plane of standard electrode positions. The nasion is facing up.

Sources. To better understand the relevance of observed non-linearity, we will now survey some sources of non-linearity that may be considered spurious.

EEG offers an example of non-linearity derived from the choice of the observable. Let us consider band-limited power and compute the shadow dataset naïvely from the sequence of power values. This would be an acceptable choice as, for example, in MEG, the computation of FC from the signal envelope is customary (19, 20). We observe a large fraction of non-linear information in all bands (Fig. 1D, yellow boxplots). However, when extracting the power from a surrogate of the original EEG time series, we notice that most of the non-linearity arises from power computation (Fig. 1D, blue boxplots).

In most bands, the non-linearity in the shadow dataset is still lower than that of the empirical one. This suggests that the transformation from voltage to power is not responsible for the entirety of the observed nonlinearity in bandpower dependence. However, the large amounts of spurious non-linearity mask the significance of band α (Fig. 1B).

As mentioned in the introduction and also discussed in detail for climate systems elsewhere (12), non-stationarities can be powerful sources of apparent non-linearity. An obvious potential source of non-linearity in the iEEG data is epileptic activity. We observe how (Fig. 4), measuring the RNL on a sliding window across the recording of a seizure, up to almost 90% of the total information appears to be due to non-linearity.

In particular, we observe how the RNL is sensitive to the signal's amplitude variation. The first peak for bands θ to γ happens when the sliding window crosses between the initial regime of low power and the first high amplitude oscillations. The RNL decreases when the window contains only large oscillations and increases again when the first group of electrodes reverts to lower power. The last increase in RNL happens when the windows cross out of the seizure.

A second example of the effect of non-stationarities comes from spiking data of individual neurons. Over the 30 minutes of recording, even without stimuli, the mice brain's activity

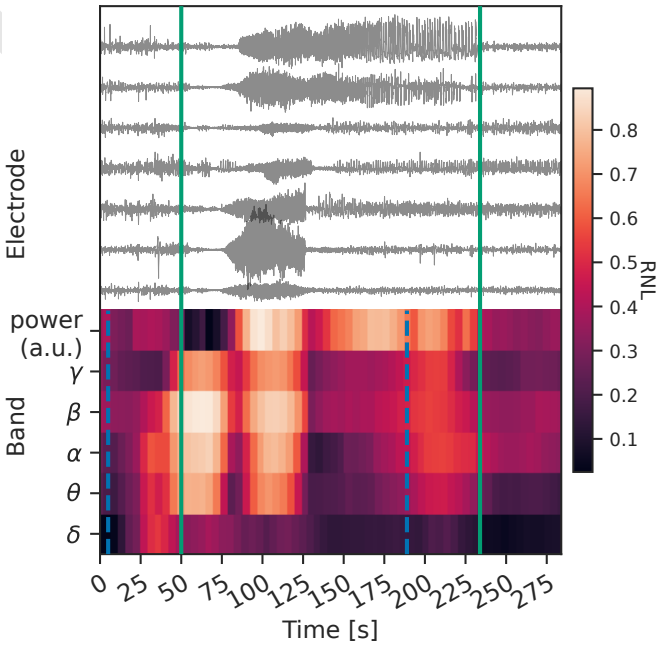

Fig. 4. Sample seizure showcasing the effect of non-stationarities on the measure of non-linearity. Top: traces from a subset of electrodes. Bottom: mowing window values of RNL and band-limited power. The green solid lines mark the beginning and the end of the seizure. The blue dashed lines mark the beginning of the first windows containing samples past the green lines.

PREPRINT

had the opportunity to drift or switch through different states (most notably, immobile and running periods). If we compute the RNL over shorter windows with fewer state changes and then average, we observe a clear reduction of the non-linearity. However, even in the last case of 3-minute epochs, the average RNL stays much above what is observed with other modalities.

Discussion

The fundamental non-linearity of neuronal activity leads to a relevant fraction of non-linear information in single-unit spikes. However, moving to more accessible and non-invasive human data, these figures are vastly reduced. This might be an effect of the scale at which the activity is recorded. Indeed, fMRI and EEG imply large spatial or temporal averages.

The more significant fraction of non-linear information in the iEEG dataset might support this interpretation. Part of this non-linearity may be explained by non-stationarities in the form of IEDs affecting the higher frequency bands and drift or state switching due to the natural activity of the subjects during the recording. However, this modality remains more promising for the detection of significant non-linearity in the human brain, and future work will have to test this with improved control over non-stationarities.

While the presence of some non-linearity is significant at all scales, its variability at the individual level makes it hard to assess its per-subject localisation, requiring group-level analysis. Indeed, the non-linearity is reduced and unreliable across sessions, which could be explained by a strong contribution of noise in the difference between the Gaussian and total MI.

Non-linearity is present at all levels of our analysis, and the oblivious use of Pearson's correlation might be inappropriate. However, we advocated caution when designing studies that plan to apply MI. Without accurate control of spurious sources of non-linearity, the improvement from using MI instead of correlation might be hardly reproducible, if any.

Lastly, many FC studies seek relatedness in the frequency domain. This practice is widespread in EEG, where there is an established relationship between frequency and function. In this context, there's a similar debate around the need to go beyond Gaussian dependencies. Future work will be able to quantitatively (beyond pure statistical testing of the binary question of presence or absence) assess the amount of non-linearity and the role of averaging in the frequency domain.

Materials and Methods

To estimate the non-linear content in the relationships between regions, electrodes, and units, we compared the Total Mutual Information (TMI) between time series to the estimate from surrogates where only the linear relationships are preserved. This allows us to evaluate the global amount of non-linearity and which are the regions or electrodes where it is more substantial.

MI estimator

Looking for the potential benefits of using MI aligns with our definition of non-linearity as the deviation from a multivariate Gaussian distribution in data. Indeed, of all distributions with Gaussian marginals, a multivariate Gaussian is maximally entropic for a given covariance matrix, i.e., has the minimum MI. Any deviation from Gaussianity will yield a higher MI. Assuming Gaussian marginal does not imply a loss of generality. Spearman correlation coefficient and MI are independent of the marginal distribution. Moreover, while approximate Gaussian distribution is often assumed, every sample distribution can be mapped to Gaussian marginals with a monotonous transformation. Indeed, we enforced Gaussian marginals via rank-normalisation to ensure precise non-Gaussianity estimates. We use the sample ranks to estimate the percentile π_i for each sample x_i and replace the original value with the one corresponding to the same percentile in the standard normal distribution $\mathcal{N}(0,1)$.

We estimate the MI of two variables through equiquantal binning (also known as equiprobable (21)): the samples are sorted on a grid with the same bin number for both variables. The bins for each variable have variable width so that the sum over the other variable always gives the same number of data points. The MI is then computed from the estimated probabilities of each bin p_{ij} as the difference between the sum of the marginal entropies of the two variables and their joint entropy:

$$MI = -\sum_i p_x^i \log p_x^i - \sum_j p_y^j \log p_y^j + \sum_{ij} p_{ij} \log p_{ij} \qquad [2]$$

with $p_x^i = \sum_j p_{ij}$, $p_y^j = \sum_i p_{ij}$, and $p_a^l \simeq p_a^k$, $a = x, y$, $\forall l, k \in [1, N]$ were the equality holds for all l and k only if the number of bins N is a divisor of the sequence length S. We chose $N = \lfloor \sqrt[3]{S} \rfloor$ in line with the previously recommended pragmatic heuristic (22).

This estimate is known to be affected by bias. For high values of MI, the estimate is bounded above by the logarithm of the number of bins. By construction, in a perfect bi-univocal relationship, the number of bins in one dimension is also the number of non-empty bins in the joint distribution, all sharing the same number of points. The higher the MI, the stronger the underestimate. On the other hand, due to the finiteness of the sample, the estimate has a positive bias (23). Any fluctuation will result in a greater than zero estimate of the MI, even for independent variables. The formulae for small sample sizes—that show good agreement with our estimates for independent variables—are derived for independent samples. However, the definition of the estimator imposes bounds on the sums of rows and columns.

These biases have opposite signs and non-trivial tractability, and we addressed them numerically. We evaluated the MI on samples from random bivariate Gaussian distributions with predetermined correlations (thus known mutual information values according to Eq. (1)) and sample size S equal to the series length. Specifically, we calculated MI for 50000 bivariate random samples of size S for each correlation value ranging from 0 to .995 in increments of .005. The average of the 50000 MI estimates approximates the expected sample mutual information for each given mutual information value. This process yields a monotonous function (with linear approximation applied if needed to ensure monotonicity for correlations near zero) that relates the true mutual information to its expected numerical MI estimate. The inverse of this function then allows us to adjust each estimated MI to produce a more accurate bias-corrected estimate of the true mutual information.

We derive the maps assuming that the number of bins N and samples S and the true MI are the only determinant factors for the bias. Every combination of a number of bins and a number of samples requires a different map. This approximation holds for a wide range of observed relations spanning over 98% of our datasets' connections.

Preliminary analyses with KNN and KDE estimators didn't show any substantial change at the price of much slower estimation. Other kinds of estimators, such as those including embedding in higher dimensional spaces (24, 25) were excluded because, even if they account for non-independent samples, they require even larger samples for a correct distribution estimate.

Non-linearity estimate. We assess non-linearity's contribution to the TMI, comparing it to the information conveyed only by linear correlations. For each dataset and subject, we calculated the MI on 99 random realisations of multivariate time series that preserve the linear structure but remove the non-linear components. If the original time series has a Gaussian dependence structure, the original data MI would be similar to that of the surrogates, up to some random error due to the variance of the estimator and

PREPRINT

surrogates. Conversely, if the original data had substantially higher MI than the surrogates, this would indicate the presence of non-linear dependencies.

We created the surrogates using the multivariate Fourier transform (FT) method (26, 27), which generates realisations of multivariate linear stochastic processes preserving the individual spectra and cross-spectra of the original time series. Specifically, each surrogate is obtained by adding identical random phases to corresponding frequency bins of the series's FT while keeping the amplitude's magnitudes unchanged. The series were then transformed back into the time domain using the inverse FT. These surrogates retain the dependency structure that a multivariate linear stochastic process can explain while destroying "non-linearity".

Comparing the MI estimate of the data and of the "linear" surrogates, rather than directly using the linear correlation of the data, has two advantages. First, correlation and MI estimators have different characteristics in terms of bias and variance, and surrogates thus allow for a direct quantitative comparison between non-linear and linear connectivity. Second, surrogates represent a suitable null model for direct statistical testing of differences. However, while these estimates are valuable for hypothesis testing—as we did to assess localisation—we use the mean of these 99 values when interested in the relative difference. We refer to this as "Gaussian" MI, which closely approximates the MI of a bivariate Gaussian distribution. The Relative Non-Linearity (RNL) is defined as the fraction of the TMI that exceeds the Gaussian MI. Note that to obtain robust estimates of RNL, throughout the paper it is reported at a global level, i.e. before taking the fraction, both TMI and Gaussian MI are first averaged over all pairs of signals.

As the correction map depends on the number of points used to estimate the MI, using oversampled data would introduce some bias. For low MI—i.e., for most of the connections—the correction map reduces the measured value. This necessary correction is smaller when the map is computed for a larger sample (i.e. lower bias). Using an oversampled sequence would lead to using a high-sample-count map, while the non-independent samples would still behave as if they were in a smaller number. As this applies to the original data and the surrogates, both measures would be overestimated by different amounts depending on their MI. This would lead to a negative bias in RNL.

We acknowledge that for fMRI data, given the band-pass filter, the current sampling rate of 0.5 Hz may lead to oversampling. However, reducing the sampling rate to 0.18 Hz would give time series too short to get any reliable estimate of MI. In this case, the relative amount of non-linear information will have a small negative bias. However, as this affects the shadow dataset (see below) similarly, every result relative to it remains valid, as will the significance tests.

Shadow datasets. Following (10), we compared each result against a control analysis using linear "shadow" datasets. This approach allows us to account for any potential biases in the generation of surrogate distributions, such as those caused by small sample sizes. For each session, a shadow dataset was created as a multivariate FT surrogate of the original, marginally normalised dataset, thereby preserving only the original data's linear structure (both autocorrelation and cross-covariance).

We then applied the same processing to the original data and the shadow, including initial normalisation, generation of multivariate surrogates, and computation of MI and RNL. This approach allowed us to replicate the entire procedure using a dataset with the same correlation and autocorrelation structure and purely linear interactions, ensuring that any potential bias in our findings due to the algorithm's numerical properties was adequately controlled. We compared the RNL between the original and shadow datasets using a paired t-test. All group-level tests applied a family-wise corrected significance threshold of $p = 0.05$.

For Band-Limited Power (BLP) data, we considered two options for generating the shadow dataset. In the naïve version, we extract the power from each band and then surrogate the power values. This procedure destroys the sequence's non-linearity and those that might have been introduced while computing the power. In the second version, we generate a surrogate voltage sequence and then extract new power values from this.

Amount of non-linearity

We relied on session-wise averages in each dataset to get the global fraction of non-linearity, RNL. In pure linear relationships, the total information measured on data should be indistinguishable from surrogates. We estimate the relative amount of non-linear information in the brain—according to the different modalities—as the relative difference between the average of the total MI across all sequence pairs and the average of the Gaussian MI across all pairs and surrogates.

Despite individual and experimental fluctuation or any systematic bias in the estimate, session-wise relative differences for linear data should stay close to the estimate from the shadow dataset. A significant difference between empirical data and shadow datasets is the sign of the presence of non-linearity.

Localisation of non-linearity

We evaluate the localisation of non-linearity leveraging the statistics of the surrogates. If a region (or, equivalently, electrode) pair has linear interaction, repeated measures across sessions or subjects should give results analogous to the surrogates. This can be observed by looking at the z-score of the measure of empirical data compared to the mean and variance of surrogate measures. For each pair of regions, we average the TMI and the 99 surrogate MIs across all subjects. We then use the mean and standard deviation of the surrogates to compute a z-score for the empirical measure. We used Holm-Bonferroni corrected p-values over empirical and shadow datasets together to assess which connections are significantly non-linear across subjects. We use z-scores instead of the relative non-linearity computed on individual pairs, as the latter introduces a strong correlation between empirical and shadow datasets due to the shared variations in correlation.

Lastly, we treated the significant z-scores as weights in an undirected graph and checked how the node degree compares with a random graph with the same link strength distribution. This allows us to visualise regions with significantly higher or lower amounts of non-linear connections compared to chance.

This analysis is possible only on the fMRI and EEG data as other modalities do not comprehensively sample the whole brain. We repeated the same steps on the fMRI data with raw preprocessing to highlight the role of artefact removal in the observed non-linearity.

Reliablity of non-linearity

For the non-linearity to be potentially relevant as an individual characteristic (e.g. as a diagnostic biomarker), the additional information provided to the researcher must be reliable across sessions. We compared how well measures of FC through correlation or TMI in one session predict FC (estimated in either way) in a different session.

We compute the strength of the connectivity in three ways: Pearson's correlation of the rank-normalised values, TMI and the Gaussian MI estimate through Pearson's correlation. The latter is obtained according to Eq. (1), with r being the sample Pearson correlation over the rank-normalised sequence. This estimate of information is faster to compute than TMI but only accounts for linear relationships as the Spearman correlation.

We assume that high or low connectivity between specific regions is a marker of scientific or clinical interest. In this case, we desire that subjects ranking high or low in one session do so also in another. For each connection, we compute the Spearman correlation of the strength of connectivity of the subjects over two sessions. We report averages across connections.

Sources of non linearity

For EEG, iEEG and single-unit spikes, we show examples of sources of non-linearity at the whole brain level. We compute the RNL on BLP sequences for EEG, contrasting two ways of deriving the

PREPRINT

shadow dataset. In iEEG data, we show the evolution of RNL throughout a seizure. We used a sliding window of 45 s with 90% overlap for each frequency band. The average power is computed as the mean over all electrodes of the absolute value of the Hilbert transform averaged over each window.

For single-unit spikes, we considered the sequences with higher sampling rates. This included firing rates averaged over windows of 125, 250 and 500 ms, allowing for enough samples to compute MI on shorter windows. We split the whole sequence into epochs of 15, 10, 6, 5 and 3 minutes, computed the RNL for each epoch and then averaged over all sessions and epochs.

Data

This work relies on five openly accessible datasets spanning four modalities: fMRI, EEG, iEEG and single-unit spikes.

fMRI data. We used two different datasets of fMRI data: the public dataset used in (28)(ESO245) and the MPI-Leipzig Mind-Brain-Body dataset (29, 30) (LEMON).

ESO245. This dataset contains 10 minutes of resting-state eyes-closed functional magnetic resonance from 245 healthy subjects (148 right-handed, 132 females, mean age 29.22/standard deviation 6.99) acquired as healthy controls as part of the ESO project. Participants were informed about the experimental procedures and provided written informed consent. The local Ethics Committee of the Institute of Clinical and Experimental Medicine and the Psychiatric Center Prague approved the study design. The acquisition included T1-weighted and T2-weighted anatomical scans not used in this study. The scanner was a 3T MRI scanner (Siemens; Magnetom Trio) at the Institute of Clinical and Experimental Medicine in Prague, Czech Republic.

Functional images were obtained using T2-weighted echo-planar imaging (EPI) with BOLD contrast. GE-EPIs (TR/TE = 2,000/30 ms) comprised 35 axial slices—acquired continuously in descending order covering the entire cerebrum (48 × 64 voxels, voxel size = 3 × 3 × 3 mm^3) (28).

Preprocessing For the present study, we used the data already processed*. We report the essential aspects of the processing while referring to (28) for the details.

The preprocessing followed the default pipeline of the CONN toolbox (McGovern Institute for Brain Research, MIT, USA) with 12 head motion parameters and five white matter components in the Component-based Correction (CompCor). The authors in (28) also detrended the resulting time series and applied band-pass filtering with cut-off frequencies 0.009-0.08 Hz. This pipeline is the "stringent" preprocessing compared to the "moderate" preprocessing (6 head motion parameters, one white matter component, cut-off frequencies 0.004-0.1 Hz) and "raw" (no CompCor nor filtering). The results in this paper use the "stringent" preprocessing unless explicitly stated otherwise.

Choice of the atlas We used two sets of atlases available in (28). The first includes 23 different parcellations using the Craddock atlas and a number of ROIs between 10 and 950 to investigate the effect of spatial averaging and region size on non-linearity. The second is the widely used AAL atlas with 90 regions to assess non-linearity localisation.

The normalised cut spectral clustering used in Craddock parcellation can yield a number of ROIs smaller than desired—i.e., empty clusters—and some ROIs can fall outside the GM mask for some subjects. This effect is more pronounced with decreasing region size and is observed in this dataset for all sizes larger than 200. The largest number of ROIs is 840 against a seed of 950. The average number of voxels included in each ROI of the Craddock atlas thus varies from almost 2×10^3 with ten regions to about 20 voxels with 840.

For each set number of ROIs in the Craddock atlas, we removed the ROIs that were empty for any subject from all subjects. We discarded the three subjects with the most empty regions to avoid discarding too many ROIs. The resulting dataset for region-size effect analysis includes 242 subjects with parcellation in 10 to 691 regions.

*Available at doi.org/10.17632/crx7d22pym.4

LEMON. We included a subset of the MPI-Leipzig Mind-Brain-Body dataset (29, 30) to assess non-linearity test-retest reliability†. The subset contains the 14 subjects (1 female, ages 20 to 35, reported in 5-year bins, mode 25-30) with at least three resting state measurement sessions, all with the same TE (to avoid possible effects due to scanning parameters). We applied the same CONN toolbox default, "stringent" preprocessing pipeline and chose the AAL 90 parcellation.

EEG data. The EEG data for this study derives from 8 minutes of resting-state eyes-closed recording from 215 healthy participants in the Max Planck Institut Leipzig Mind-Brain-Body Dataset – LEMON. We report here the main features of the data referring to the dataset presentation papers (29, 30).

Data acquisition. The EEG data was acquired with a sampling frequency of 2500 Hz using 62 channels (ActiCAP, Brain Products GmbH, Gilching, Germany) according to the 10-10 system with one VEOG electrode. The total acquisition lasted for 16 minutes, divided into 60 s blocks with 8 eyes closed blocks interleaved with 8 eyes open blocks.

Data preprocessing. We downloaded the preprocessed version of the dataset containing 204 subjects‡. The preprocessing included bandpass filtering (1–45 Hz), downsampling to 250 Hz, removal of corrupted channels, and eye movement and heartbeat removal via ICA. We excluded from the analysis 7 electrodes frequently missing (T7, T8, Cz, F7, CP6, PO10, Fp2) and considered the 150 subjects with all the remaining 54 channels available.

We further processed the data in Python using the MNE (31) and SciPy (32) packages. At this stage, the sequences are split into segments up to 60 s long. We obtained the five usual frequency bands ($\delta = [1, 4]$ Hz, $\theta = [4, 8]$ Hz, $\alpha = [8, 12]$ Hz, $\beta = [12, 30]$ Hz, $\gamma = [30, 44]$ Hz) from each segment using an IIR Butterworth filter of order 4 in two forward and backwards passes to minimise phase distortion. We obtained the scalp voltage by down-sampling to 1.25 times the Nyquist frequency of the upper limit of each band of the filtered sequence to avoid biases (see "MI estimator").

We derived the scalp BLP from the filtered signal by applying Hilbert transformation and taking block averages of the modulus over 125 ms windows. We obtained the shadow dataset for BLP sequences computing FT multivariate surrogates before applying the Hilbert transformation.

Finally, we obtained the three epochs used for RNL and reliability estimation by chaining together segments for a total of 124 s. We obtained the naïve shadow dataset as a surrogate of the BLP sequences right before RNL estimation.

iEEG data. The iEEG dataset is derived from the open dataset published by the Sleep-Wake-Epilepsy-Center (SWEC) of the University Department of Neurology at the Inselspital Bern and the Integrated Systems Laboratory of the ETH Zurich (33). The original dataset§, contains 2656 hours of anonymised and continuous intracranial electroencephalography (iEEG) of 18 patients with pharmaco-resistant epilepsies. All the patients gave written informed consent that their iEEG data might be used for research and teaching purposes. Given the extensive size of this dataset, for each subject, we selected the central 124 s of the first hour that was at least 45 minutes away from any seizure based on metadata and manual inspection.

Data acquisition. The iEEG signals were recorded intracranially by strip, grid, and depth electrodes. The recorded signal was saved at a rate of 512 or 1024 Hz after band-passing between 0.5 and 120 Hz with a double-pass fourth-order Butterworth filter. An epileptologist visually inspected all the recordings, marked the onset and end of each seizure, and removed corrupted channels.

Data preprocessing. The 18 subjects have between 24 and 128 electrodes of undisclosed type in undisclosed locations. Also, it is

†Available at: https://fcon_1000.projects.nitrc.org/indi/retro/MPI_LEMON/downloads/download_MRI.html

‡Available at: https://fcon_1000.projects.nitrc.org/indi/retro/MPI_LEMON/downloads/download_EEG.html

§Available at: http://ieeg-swez.ethz.ch/

PREPRINT

not reported which electrodes are located in the epileptic foci. To improve data homogeneity, we randomly selected 24 electrodes for each subject. We band-passed and down-sampled the time series as we did for the EEG electrode voltage and selected a segment of 124 s. We extracted three additional subsets of the data. In the first, the sequences have the same starting time but cover different time lengths (up to 22 minutes and 44 s) depending on the frequency band to allow in each band the same number of samples as for band gamma with 124 s. The other two correspond to windows of 124 s extracted 24 and 48 hours after the first one. In case of a seizure or a total length of the recording under 48 hours, we selected seizure-free hours, trying to keep the distance between windows as close as possible to 24 hours.

Single Unit Spikes data. The last dataset we include is derived from the Allen Brain Observatory – Neuropixels Visual Coding dataset (34). The dataset[¶] contains individual units isolated in 58 mice from six probes inserted in the visual areas. The dataset section relevant to this work includes spiking times from individual units and metadata about unit localisation and quality. We considered the 26 specimens that underwent the "functional connectivity" experiment. This included 30 minutes of spontaneous activity, i.e., resting-state data.

Data preprocessing. With this dataset, we aimed to explore the effects of averaging on non-linearity with the finest granularity. We first selected units with reasonable quality measures (expected fraction of missed spikes ≤ 0.08, maximum inter-spike interval violations $= 0.2$ (35)). Then, we created groups of 32 units from the same brain structure. If a structure had enough units to form more than one group, we determined the groups so that the units were spatially separated. When the number of exceeding units is not enough to form an extra group, we discarded those of lower quality weighting in order: ISI violations, quality if isolation from other units, and the expected fraction of missed spikes. We kept only the 16 sessions that allowed the identification of at least nine groups.

In this case, our data will be the average spike count per unit in time intervals of different lengths. As our estimator is designed to work with continuous variables, we added a slight jitter to the data to keep all points distinct. For every session, we derived 42 sets of time series varying the time interval from 125 ms to 8 s (doubling the length at each step) and taking the average spike count of 1 to 32 units (doubling the number at each step), assessing thus the effect of either temporal or spatial averaging.

ACKNOWLEDGMENTS. The publication was supported by ERDF-Project Brain dynamics, No. CZ.02.01.01/00/22_008/0004643, the Czech Science Foundation projects No. 21-32608S and No. 21-17211S.

1. KJ Friston, Functional and effective connectivity in neuroimaging: A synthesis. *Hum. Brain Mapp.* **2**, 56–78 (1994).
2. KJ Friston, Functional and Effective Connectivity: A Review. *Brain Connect.* **1**, 13–36 (2011).
3. ME Raichle, et al., A default mode of brain function. *Proc. Natl. Acad. Sci. United States Am.* **98**, 676–682 (2001).
4. BB Biswal, et al., Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. United States Am.* **107**, 4734–4739 (2010).
5. H Bakhshayesh, SP Fitzgibbon, AS Janani, TS Grummett, KJ Pope, Detecting synchrony in EEG: A comparative study of functional connectivity measures. *Comput. Biol. Medicine* **105**, 1–15 (2019) application to EEG, good grades to MI with EQB.
6. AS Mahadevan, UA Tooley, MA Bertolero, AP Mackey, DS Bassett, Evaluating the sensitivity of functional connectivity measures to motion artifact in resting-state fmri data. *NeuroImage* **241**, 118408 (2021).
7. Z Wang, A Alahmadi, D Zhu, T Li, Brain functional connectivity analysis using mutual information in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. (IEEE), pp. 542–546 (2015).
8. H Chen, Y Song, X Li, A deep learning framework for identifying children with ADHD using an EEG-based brain network. *Neurocomputing* **356**, 83–96 (2019).
9. HE Wang, et al., A systematic framework for functional connectivity measures. *Front. Neurosci.* **8**, 111632 (2014).
10. J Hlinka, M Paluš, M Vejmelka, D Mantini, M Corbetta, Functional connectivity in resting-state fmri: Is linear correlation sufficient? *NeuroImage* **54**, 2218–2225 (2011).
11. N Tzourio-Mazoyer, et al., Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15**, 273–289 (2002).
12. J Hlinka, D Hartman, M Vejmelka, D Novotná, M Paluš, Non-linear dependence and teleconnections in climate data: Sources, relevance, nonstationarity. *Clim. Dyn.* **42**, 1873–1886 (2014).
13. D Hartman, J Hlinka, Nonlinearity in stock networks. *Chaos* **28** (2018).
14. C Chang, GH Glover, Time–frequency dynamics of resting-state brain connectivity measured with fmri. *NeuroImage* **50**, 81–98 (2010).
15. J Cabral, ML Kringelbach, G Deco, Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *NeuroImage* **160**, 84–96 (2017).
16. RC Craddock, GA James, PE Holtzheimer, XP Hu, HS Mayberg, A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* **33**, 1914–1928 (2012).
17. SM Motlaghian, et al., Nonlinear functional network connectivity in resting functional magnetic resonance imaging data. *Hum. Brain Mapp.* **43**, 4556–4566 (2022).
18. B Frauscher, et al., Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas. *Brain* **141**, 1130–1144 (2018).
19. MJ Brookes, et al., Measuring functional connectivity using MEG: Methodology and comparison with fcMRI. *NeuroImage* **56**, 1082–1104 (2011).
20. JF Hipp, M Siegel, Bold fmri correlation reflects frequency-specific neuronal correlation. *Curr. Biol.* **25**, 1368–1374 (2015).
21. GA Darbellay, I Vajda, Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Inf. Theory* **45**, 1315–1321 (1999) equiprobable intervals.
22. M Paluš, Testing for nonlinearity using redundancies: quantitative and qualitative aspects. *Phys. D* **80**, 186–205 (1995).
23. JA Bonachela, H Hinrichsen, MA Muñoz, Entropy estimates of small data sets. *J. Phys. A: Math. Theor.* **41**, 202001 (2008).
24. RQ Quiroga, A Kraskov, T Kreuz, P Grassberger, Performance of different synchronization measures in real data: A case study on electroencephalographic signals. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **65**, 14 (2002).
25. Z Jia, Y Lin, Y Liu, Z Jiao, J Wang, Refined nonuniform embedding for coupling detection in multivariate time series. *Phys. Rev. E* **101**, 062113 (2020).
26. D Prichard, J Theiler, Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.* **73**, 951 (1994).
27. M Paluš, Detecting phase synchronization in noisy systems. *Phys. Lett. A* **235**, 341–351 (1997).
28. J Kopal, A Pidnebesna, D Tomeček, J Tintěra, J Hlinka, Typicality of functional connectivity robustly captures motion artifacts in rs-fMRI across datasets, atlases, and preprocessing pipelines. *Hum. Brain Mapp.* **41**, 5325–5340 (2020).
29. A Babayan, et al., A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Sci. Data 2019 6:1* **6**, 1–21 (2019).
30. N Mendes, et al., A functional connectome phenotyping dataset including cognitive state and personality measures. *Sci. Data 2019 6:1* **6**, 1–19 (2019).
31. A Gramfort, et al., MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 1–13 (2013).
32. P Virtanen, et al., SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17**, 261–272 (2020).
33. A Burrello, L Cavigelli, K Schindler, L Benini, A Rahimi, Laelaps: An Energy-Efficient Seizure Detection Algorithm from Long-term Human iEEG Recordings without False Alarms in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. (IEEE), pp. 752–757 (2019).
34. JH Siegle, et al., Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature* **592**, 86–92 (2021).
35. DN Hill, SB Mehta, D Kleinfeld, Quality metrics to accompany spike sorting of extracellular signals. *J. Neurosci.* **31**, 8699–8705 (2011).

[¶] Available at: https://allensdk.readthedocs.io/en/latest/visual_coding_neuropixels.html

PREPRINT